

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 14:47:29 ; Search time 2236 Seconds  
(without alignments)  
17303.675 Million cell updates/sec

Title: US-08-961-083-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGACTGT.....TAAGTAAGCAAAATAAAC 2389

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estchum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.8	3.0	906	17	BH153606 ENTTS83TF
2	69.8	2.9	890	17	BH146886 ENTPK48TF
3	67.8	2.8	180	17	AF325819 AF325819
4	67.8	2.8	908	17	A2548467 A2548467
5	64.8	2.7	843	17	A2551618 ENTDSV4TR
6	63.8	2.7	877	17	A2531291 ENTB034TR

7	62.2	2.6	912	17	A2551092	A2551092 ENTFCJ22TF
C	60.8	2.5	849	17	A2546039	A2546039 ENTPW53TF
8	59.2	2.5	931	17	BH160272	BH160272 ENTQV49TR
C	58.6	2.5	1101	17	CNS00396	A1063821 Drosoph11
10	57.2	2.4	900	17	A2549980	A2549980 ENTDD94TF
C	56.8	2.4	905	17	A2550256	A2550256 ENTBV58TR
12	56.6	2.4	645	13	BH165350	BH165350 EST567873
13	56.6	2.4	747	13	BH162732	BH162732 EST565255
C	55.2	2.3	816	17	A2535744	A2535744 ENTQ025TR
15	55.2	2.3	647	17	BH163120	BH163120 EST565543
16	55	2.3	671	13	BH160252	BH160252 EST562175
17	55	2.3	694	13	BH168242	BH168242 EST570765
18	55	2.3	717	13	BH160500	BH160500 EST563023
19	55	2.3	756	13	BH162492	BH162492 EST565015
20	55	2.3	794	13	BH159906	BH159906 EST562429
21	55	2.3	795	17	A2528485	A2528485 ENTCM64TF
22	55	2.3	823	17	A2676218	A2676218 ENTKE36TR
C	55	2.3	843	17	BH139532	BH139532 ENTNG88TF
23	55	2.3	890	17	A2530768	A2530768 ENTBH54TF
24	52.8	2.2	677	13	BH161314	BH161314 EST563837
25	52.4	2.2	725	13	BH169486	BH169486 EST572009
26	52.4	2.2	866	17	A2527885	A2527885 ENTCA79TR
27	52.4	2.2	976	17	BH149983	BH149983 ENTQD93TF
28	52.4	2.2	945	17	CNS008CA	AL051492 Drosoph11
29	52.2	2.2	435	13	BH169710	BH169710 EST572233
30	51.8	2.2	880	17	A2529191	A2529191 ENTBV68TR
C	51.6	2.2	853	17	CNS023KH	A179594 Tetraodon
32	51.6	2.2	500	9	AU088119	AU088119 AU088119
33	51	2.1	726	17	A2196050	A2196050 SP_1031_A
34	50.8	2.1	1007	17	CNS06X9S	AL19462 T3 end of
C	50.2	2.1	445	13	BH167469	BH167469 EST569592
35	50	2.1	707	13	BH164483	BH164483 EST567006
36	49.8	2.1	598	13	BH170666	BH170666 EST573189
C	49.6	2.1	605	13	BH163520	BH163520 EST566043
37	49.6	2.1	629	13	B073162	B073162 B073162
38	49.6	2.1	687	13	BH169665	BH169665 EST572188
39	49.4	2.1	1101	17	CNS0022U	AL097152 Drosoph11
40	48.8	2.0	605	13	BH162501	BH162501 EST565024
41	48.6	2.0	891	17	A2683582	A2683582 ENTRK47TR
42	48.6	2.0				
43	48.6	2.0				
44	48.6	2.0				
45	48.4	2.0				

#### ALIGNMENTS

RESULT 1  
BH153606/c 906 bp DNA linear GSS 24-SEP-2001  
LOCUS ENTTS83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.  
ACCESSION BH153606  
VERSION BH153606.1 GI:15725323  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 906) Eukaryota: Entamoebidae; Entamoeba.  
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library (2001)  
COMMENT Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@igf.org

DNA library  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
Seq primer: M13-Forward  
Class: Shotgun  
High quality sequence start: 15  
High quality sequence stop: 733.



Oy	1737	TACCTCATTCTCCACAGCAGATGTTAAAGCAATTCACACTGGAAATATGTGCAGCAGCTA	1796
Db	255	AAGACGATGATGAAAGAACGCATGTATGAAAGAACGATGATGAAAGAACGATGTAAG	314
Oy	1797	TTTCAATCGTGTGAAAGGGGAAACGAAATTCACCTGGTTCGACTTCATATATGTTG	1856
Db	315	AAGACGATGATGAAAGAACGCATGTATGAAAGAACGATGATGAAAGAACGATGTAAG	374
Oy	1857	AGCATACAGTTGAGGTTAAAAAGGTAATTGTTGATTATTCCTATAGAATCATTTACATA	1916
Db	375	AAGACGATGATGAAAGAACGCATGTATGAAAGAACGATGATGAAAGAACGATGTAAG	434
Oy	1917	ATATTAAATTTGCTGTGGTTTGATGATCAACATCAAAAGCTCCAATGGCTTACTCTTG	1976
Db	435	AAGACGATGATGAAAGAACGCATGTATGAAAGAACGATGATGAAAGAACGATGTAAG	494
Oy	1977	AAGATTTGTTTGCAGCATTAAGTACTACGTAGAAACCCCTGACGAACTCCACATTTCTA	2036
Db	495	AAGATGAAGATGATGAAAGAACGCATGTATGTAAGAACGATGATGTAAGAACG	554
Oy	2037	ATGATGATGGGGCAATGCCAGTGCATCGTGTGAGCAAGAAAGAACCACTGTAAGATGC	2096
Db	555	ACGATGAAGACGACGATGAAGACGCAGATGAAGACGATGAAGAAAGCGATGTAAGAG	614
Oy	2097	CAAAATGAACCTTCAAAACCGATGTAAGACCCAGTAGAGAGGAAAAACACTGCTGCGCCAGAG	2156
Db	615	ACGATGAAGAACGATGAAGAAAGCAGATGAAAGAACGATGAAGAAAGCAGATGTAAGAG	674
Oy	2157	TGCCCAAGTGAAGCTGAAAAAGTGAAGACCCAACTCAAGAACGACAGATT	2209
Db	675	ACGATGAAGAACGATGAAGAACGCATGTAAGAACGATGATGAAGAACGAT	727

RESULT <sup>1</sup> 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	SOURCE
AF325819	AF325819	Leptospira interrogans serovar lai genomic clone 18, DNA sequence.	AF325819	AF325819	GSS.	GI:19263459	180 bp	DNA	linear	GSS 08-MAR-2002			/organism="Leptospira interrogans serovar lai" /strain="017" /db_xref="taxon:57678" /clone="18" /clone_lib="Leptospira interrogans serovar lai DNA" /note="obtained through DNA subtraction using the nonpathogenic strain Leptospira biflexa serovar Patoc strain Patoc I"	location/qualifiers 1..180
													Leptospira interrogans serovar lai. Leptospira interrogans serovar lai Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira. 1 (bases 1 to 180) Hu,C. and Bao,L. Leptospira interrogans serovar lai DNA Unpublished (2002) Contact: Hu C Leptospiriosis Research Unit West China University of Medical Sciences 17# Renming Nan Road, Chengdu, Sichuan 610041, P.R. China Email: huchanghua@263.net Class: unknown.	
														BASE COUNT 66 a 27 c 40 g 47 t ORIGIN

	Query Match	2.8%	Score 67.8;	DB 17;	Length 180;
	Best Local Similarity	62.9%	Pred. No. 3.6e-06;		
	Matches 105; Conservative	0;	Mismatches 62;	Indels 0;	Gaps 0;
QY	249 TACTATGAAAGTCCAACTTATAAGCTAAAAAGATGACGATTTTGTTAATACGGTCMAAG	308			

Db 1 TATATATACGACGCCCAATATACGTTTAAACATCTGATGTTTAAACGAGATTCTAG 60

QY GTGCGATATGTTATCAAGTAGTAGTGGAAAACTATAGTTTACCTTTAAGATGTCTGCCACG 368

Db 61 ATGGCTATATATTTATCAAGGTTGATGCGCAAGTATTTATGTTTATTTAAACCGCAGAGCAAGC 120

QY CGGATTAACGTCCTGTCAAAAGAGAAATCAATCAATCGACAAAACACAGAG 415

Db 121 GTAAAAAATATCCGTACCAAGCGAGCAAAATTTGCTGAAACAGGTTGAAAAAG 167

RESULT -4				
AZ548467				
LOCUS	AZ548467	908 bp	DNA	linear GSS 14-NOV-2000
DEFINITION				
	ENTEXK30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic , DNA sequence.			
ACCESSION	AZ548467			
VERSION	AZ548467			
KEYWORDS	AZ548467.1 GI:11172102			
SOURCE	GSS.			
ORGANISM	Entamoeba histolytica.			
	Entamoeba histolytica.			
	Eukaryota; Eutamoebidae; Entamoeba.			
REFERENCE	1 (bases 1 to 908)			
AUTHORS	Lofthus,B., Van Aken,S. and Fraser,C.			
TITLE	Determination of clone end sequences from Entamoeba histolytica HM:IWMS sheared DNA library			
JOURNAL	unpublished (2000)			
COMMENT	Contact: Brendan J Lofthus			

Email: [bjloftus@iclr.org](mailto:bjloftus@iclr.org)  
 Clones are derived from the *Entamoeba histolytica* HMI:IM5 sheared  
 DNA library  
 Seq primer: M13-Reverse  
 Class: shotgun  
 High quality sequence start: 17  
 High quality sequence stop: 828.

FEATURES	SOURCE
Location/Qualifiers	1. 908
/organism="Entamoeba histolytica"	
/strain="HM1:IMSS"	
/db_xref="taxon:5759"	
/clone_lib="Entamoeba histolytica Sheared DNA"	
/note="Vector: pHOSt1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450-.) The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."	434 a
	16 c 289 g 169 t

	Query Match	Score	67.8	DB	17.8	Length	908;			
	Best Local Similarity	44.0%	Pred.	6.2e-05;						
	Matches	334;	Conservative	0;	Mismatches	422;	Indels	3;	Gaps	1;
OY	1555	TGAGTATCTGTGAACGAGATTTCGTATGGCTCCTCAATTAGCTGTAATAATPACAACTCAGA	1584							
Db	33	TGAAGATATGATGATGTGAAGAAGATGATGAGAACMAAGATCGATGTGAAGAAGATGATGA	92							
OY	1585	TGGTTACATTTTTTGATGACAATGATATATATATAGTCATAGAGAGATGCATATGTAAACC	1644							
Bb	93	TGATGACAAAGATGATGTGAAGAAGATGAGAAGAACMAAGATCGATGTGAAGAAGATGA	152							

1645 TCATATGGCCATAGTCACTGATG---GAAAGATAGCCTTTCGATTAAGGAAAGT 1701  
153 TGAAGACGATGAAGACGACCAATATGATTTAGAGATGATGATGATGAAGACAGAGA 212  
1702 TGCAGCTCAAGCCTATTAAGAAAAAGGTATCTCACTCCATCCAGACGCAGATGT 1761  
213 TGATGATGAAGATGATGAAGACGATGAGACGAAATATGAATTAAGATGATGATGA 272  
1762 TAAGCAAAATCCAACTGGAGATAGTGCAGCCTATTATCAATTCGTCGTAAGGCGAAA 1821  
273 TGATGAAGAAGAAAGATGATGATGAAGAAGAAAGATGATGAAGATGATGATGAAGA 332  
1822 ACAATTCCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1881  
333 AGAAGATGAAGAAGATGATGAAGAAGATGATGAAGATGATGATGAAGAAGATGA 392  
1882 TAATTTGATTTATCTCAATTAAGATCAATTAATTAATTAATTTGCTTGGTTGATGA 1941  
393 TGATGAAGATGATGAAGACGATGAAGACGAAATATGAATTAAGATGATGATGAAGA 452  
1942 TCACACATTAAGATCCCAATGGCTATACCTTGGAAGATTTGTTCCGACATTAAGTA 2001  
453 AGAAGAAGATGATGATGAAGAAGAAGATGATGATGAAGATGATGAAGAAGAAGA 512  
2002 CTACGTAGAACACCCCTGACGACGATCCACATTTCTAATGATGATGGGCAATGCCAGTGA 2061  
513 TGAAGATGATGATGAAGAAGATGATGAAGATGATGATGAAGAAGATGATGATGAAGA 572  
2062 GCATGTTTATGAGCAAGAAAGACCAAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGA 2121  
573 AGAAGATGAAGAAGAAAGATGATGAAGAAGATGATGAAGAAGAAAGATGATGAAGA 632  
2122 AGAGCCAGTGAAGAAGAACCTGCTGAGCCAGAAAGTCCCACTGATGAGCTGAAAAAGT 2181  
633 AGACGACGATGATGAATTAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGA 692  
2182 AGAAGCCCACTCAAGAAAGACAGAAAGTTTGTTCGCAAGATGAAGGATCTAGTCTGAA 2241  
693 AGATGATGAAGAAGAAAGATGATGATGATGATGATGATGAAGAAGATGATGATGAAGA 752  
2242 AGCCAAATGCACAGAAACTCTAGCTGTTTACGAATTA 2280  
753 TGAAGAAGATGAAGATGATGATGATGAAGAAGATGA 791

RESULT 5  
LOCUS A2551618/c 843 bp DNA linear GSS 14-NOV-2000  
DEFINITION ENT549R Entamoeba histolytica sheared DNA Entamoeba histolytica  
ACCESSION A2551618  
VERSION A2551618.1 GI:11176919  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 843)  
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HML:IMSS sheared DNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@igr.org  
Clones are derived from the Entamoeba histolytica HML:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun

High quality sequence start: 39  
High quality sequence stop: 838.  
Location/Qualifiers  
1..843  
/organism="Entamoeba histolytica"  
/strain="HML:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica sheared DNA"  
/note="Vector: PHOS1, Site 1; Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + l method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaubin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 155 a 266 c 18 g 404 t  
ORIGIN

Query Match 2.7%; Score 64.8; DB 17; Length 843;  
Best Local Similarity 44.1%; Pred. No. 3.4e-05;  
Matches 319; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

QY 1521 AAATTGATTAAGTGAAGACGAAAGTTCCTCAATTAAGTGAATTAAGTGAATTAAGTGA 1580  
DB 753 AAGAAGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAG 694  
QY 1581 CAGATGTTTACAT---TTTGATGAACATGATTAATCAAGTGAAGATGATGATGATGATG 1637  
DB 693 AAGATGATGATGATGAAGAAGAAAGATGATGATGAAGATGATGAAGATGATGAAGACG 634  
QY 1638 TAAGCCCTCATATGAGCCATAGTCACTGATGATGAAGAAGATTAAGTGAATTAAGTGA 1697  
DB 633 AGCAATATGATTAAGAAAGATGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGA 574  
QY 1698 AAGTTGACCTCAAGCCTTACTTAAGAAAGATTTCTCACTCCATCTCCAGACGACAG 1757  
DB 573 AAGATGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAG 514  
QY 1758 AAGTTAAGCAATTCACACGAGATGATGATGATGATGATGATGATGATGATGATGATG 1817  
DB 513 AAGATGATGATGATGAAGAAGAAAGATGATGATGAAGAAGATGATGAAGAAGATGATG 454  
QY 1818 AAAAAGAAATTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1877  
DB 453 AATATGAATTAAGAAAGATGATGAAGAAGAAAGATGATGAAGAAGATGATGAAGAAG 394  
QY 1878 AAGCTAATTTGATTTATCCCTAATAGATCATTAATTAATTAATTTGCTGTTGG 1937  
DB 393 ATGATGAAGATGATGATGAAGAAGAAAGATGATGATGAAGAAGATGATGAAGAAGATG 334  
QY 1938 ATGATCAACATTCAAAGCTCCAAATGGCTATATCTTGAAGATTTGTTGTCAGACATTA 1997  
DB 333 ATGATGAAGAAAGATGATGATGAAGAAGAAAGATGATGAAGAAGATGATGAAGAAG 274  
QY 1998 AGTACTACGTAGAACACCTGACGACGATCCATTTCTATGATGATGATGATGATGATG 2057  
DB 273 ATGATGAAGAAGAAAGATGATGATGAAGAAGAAAGATGATGAAGAAGATGATGAAGA 214  
QY 2058 GTGACATGTTGTTAGCAAGAAAGACACAGTGAAGATCCAAATTAAGATTTCAAGCG 2117  
DB 213 AAGATGATGAAGAAGATGATGAAGAAGAAAGATGATGAAGAAGAAAGATGATGAAGA 154  
QY 2118 ATGAAGACGATGAAGAAAGACCTGCTGAGCCAAAGTCCCTCAAGTGAAGACGATGA 2177  
DB 153 ATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGA 94  
QY 2178 AAGTAGAACCCCACTCAAGAAAGATTTGCTTGCAGAAAGTAAAGATTTAGTTC 2237







```

Db      442 GATGATGATTTGAATTAGAAAGATGAAAGATGAAAGATGAAAGATGATGATCAT 383
Oy      1550 ATTGCTCATTTAGCGTATAGATATACAGTCAGATGCTTACATTTTGTATCAATGAT 1609
Db      382 GATGATGAAAGATGATGAAAGATGATGAAAGATGATGAAAGATGATGAAAGATGAT 323
Oy      1610 ATATATGATGATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
Db      322 GAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 263
Oy      1670 GAAAAAGATAGGCTTTCTGATTAAGAAAAAGTTGCACTCAAGCTTATPCTAAGAAAA 1729
Db      262 GAAAGATGAAAGATGAAAGATGATGATGAAAGATGATGAAAGATGATGATGATGAT 203
Oy      1730 GGATATCTTCCCTCCATCTCCACAGCAGATGTTAAAGCAATCCCACTGGAATAGTGA 1789
Db      202 GAATTAAGAAAGATGAAAGATGATGAAAGATGAAAGATGATGAAAGATGATGAAAGAA 143
Oy      1790 GCAGCTATTACATCGTGTGAAGGGGAAAAA 1822
Db      142 GAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 110

```

```

RESULT 9
LOCUS   BH160272
DEFINITION  ENTQV49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
          genomic, DNA sequence.
ACCESSION  BH160272
VERSION    BH160272.1 GI:15733710
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
REFERENCE  1 (bases 1 to 931)
AUTHORS    Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
           HMI:IMSS sheared DNA library (2001)
           Unpublished (2001)
JOURNAL    Contact: Brendan J Loftus
           Department of Eukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0208
           Fax: 301 838 3543
           Email: b.loftus@iglr.org
           Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
           DNA library
           Seq primer: M13-Reverse
           Class: shotgun
           High quality sequence start: 8
           High quality sequence stop: 677.
           Location/Qualifiers
               1..931
                   /organism="Entamoeba histolytica"
                   /strain="HMI:IMSS"
                   /db_xref="taxon:5759"
                   /clone_lib="Entamoeba histolytica Sheared DNA"
                   /note="Vector: PHOSI; Site: 1; Bst I; Constructed at the
                   Institute for Genomic Research (TIGR), Rockville, MD.
                   Genomic DNA isolated from broth cultures of E. histolytica
                   using a method described by Clark and Diamond (Clark,
                   C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                   method for isolate identification. Exp. Parasitol.
                   77:450.). The DNA was mechanically sheared to give a
                   tight size distribution (~2 kb). The v + i method used for
                   the library construction is described in detail in Smith,
                   H.O. and Venter, J.C. (Making small insert libraries for
                   whole genome shotgun sequencing projects. In Genome
                   Sequencing: A Practical Approach, eds. M. Vaubin and B.
                   Barrell, Oxford University Press, 1999)."
```

BASE COUNT

402 a 51 c 279 g 199 t

## ORIGIN

Query Match 2.5%; Score 59.2; DB 17; Length 931;  
 Best Local Similarity 43.3%; Pred. No. 0.00086;  
 Matches 277; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```

Oy      1583 GATGCTTACATTTTGTATGATCAATCATGATATATATCACTGATGAAGAGATGCATATGTAACG 1642
Db      17 GATGATGATGAAGAAAGAAAGATGATGATGAAGATGATGAAGATGATGAAGATGATGAAG 76
Oy      1643 CCTCATATGGGCCATAGTCCTGATTTGAAAAGATAGCCTTTCTTGATTAAGAAAAAGTT 1702
Db      77 TATGATTTAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 136
Oy      1703 GCAGCTCAAGCCTATATCTAAAGAAAAAGTATCCATACCTCCATCTCCAGACCCAGATGTT 1762
Db      137 GATGATGAAGATGATGATGAAGAAAGATGAAAGATGATGAAGAAAGATGATGAAGAT 196
Oy      1763 AAAGCAATCCAACTGAGATAGTCAGCAGCAGCTATTTTCAATCGTGTGAAGGGGAAAAA 1822
Db      197 GATGATGAAGAAAGAAAGATGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGAT 256
Oy      1823 CGAATTCACCTGCTTCGACTTCATATATGTTGACATACACTTGAGGTTAAAAACGCT 1882
Db      257 GAATTAAGAAAGATGATGATGAAGAAAGATGATGATGATGAAGAAAGATGATGATGATGAT 316
Oy      1883 AATTTGATATATTCCTCATAGATCATTTCCATATATTAATTTGCTGTTGATGAT 1942
Db      317 GAAATGATGATGAAGAAAGAAAGATGATGATGAAGAAAGATGATGAAGATGATGATGAT 376
Oy      1943 CACACATACAAAGCTCCAAATGGCTATACCTTGAAGATTTGTTGCGACGATTAAGTAC 2002
Db      377 GAAAGAAAGATGATGATGAAGAAAGATGAAGAAAGATGAAGAAAGATGATGAAGATGAT 436
Oy      2003 TACGTAGAACACCTTGACGACAGCTCCACATTTCTATGATGATGGGCAATCCAGTGA 2062
Db      437 GAAAGAAAGAAAGATGATGATGAAGAAAGATGATGAATTTGAAGATGATGAAGAAAG 496
Oy      2063 CATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAATTAACCTTCAACCGGATGAA 2122
Db      497 GATGAAGAAAGATGATGAAGAAAGATGATGAAGAAAGATGATGATGATGATGATGAT 556
Oy      2123 GAGCCAGTAGAGAAACACCTGCTGAGCCGAGAGTCCCTCAAGTAGAGAGATGAAGAAATGA 2182
Db      557 GAAAGAAAGATGATGAAGATGATGAAGAAAGATGAAGATGATGAAGAAAGATGAAGATGAT 616
Oy      2183 GAGCCCAACTCAAGAAAGCAGAGATTTTCTTGGCAAG 2222
Db      617 GAAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAG 656

```

```

RESULT 10
LOCUS   CNS00396G
DEFINITION  Drosophila melanogaster genome survey sequence TE73 end of BAC #
           BACR08K10 of RPCT-98 library from Drosophila melanogaster (fruit
           fly), genomic survey sequence.
ACCESSION  AL063921
VERSION    AL063921.1 GI:4941778
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster.
ORGANISM   Drosophila melanogaster.
           Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
           Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
           Ephydroidae; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
AUTHORS    Genoscope.
TITLE       Direct Submission
JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
           BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
           - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
           collaboration with the Berkeley Drosophila Genome Project (BDGP).
```

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.flytily.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mosmoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

**source**

1. .1101

```
/organism="Drosophila melanogaster"
```

```
/db_xref="taxon:7227"
```

```
/clone="BACR08K10"
```

```
/clone_lib="RPCI-98"
```

```
/note="end : TEXT3"
```

BASE COUNT	201 a	64 c	131 g	202 t	503 others
------------	-------	------	-------	-------	------------

## ORIGIN

Query Match

Best Local Similarity 16.2%; Pred. No. 0.0013;

Matches 97; Conservative 252; Mismatches 251; Indels 0; Gaps 0;

## FEATURES

## Source

1. .900

/organism="Entamoeba histolytica"

```
/strain="HML:IMSS"
```

/db\_xref="taxon:5759"

```
/clone_lib="Entamoeba histolytica Sheared DNA"
```

/note="Vector: PHOS1; site\_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

BASE COUNT	376 a	63 c	235 g	226 t
ORIGIN				

## ORIGIN

Query Match	2.48; Score 57.2; DB 17; Length 900
-------------	-------------------------------------

Best Local Similarity 45.58; Pred. No. 0.0027;

Matches	203;	Conservative	0;	Mismatches	243;	Indels	0;	Gaps	0
---------	------	--------------	----	------------	------	--------	----	------	---

186 ATGGCGACCACCTATCATTTATACATGGTAAGGTTCCTTATGACGCTATCATCAGTGAAG 245

[illegible][illegible][illegible]

246 AATTACTCATGAGATCCAACTATAAGCTAAAAGATGAGGATATTGTTAATGAGGTCA 305

140  
130  
120  
110  
100  
90  
80  
70  
60  
50  
40  
30  
20  
10  
0

Db 440 ATGAAGAAGATGAAGATGAAGACCGATGATGAAGATGATGAAGATGATGAAG 499

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

306 AGGGTGATATGTTATCAAGGTAGATGGAATACTACTATGTTACCTTAAGGATGCTGCC 365

Db 500 ATGATGAGATGATGAAGATGATGAAGATGATGAAGATGATG 559

LOCUS	AZ549980	900 bp	DNA	linear	GSS 14-NOV-2000
DEFINITION	ENTID94Tf Entamoeba histolytica Sheared DNA			Entamoeba histolytica genomic, DNA sequence.	





1845 CATATATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTGATTATTCCTCATAAGG 1904

Job time : 2262 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 14:47:29 ; Search time 71 Seconds  
(without alignments)  
10319.025 Million cell updates/sec

Title: US-08-961-083-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAGAATAAATAAAC 2389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PCtus\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2388	100.0	2389	US-08-961-083-55	Sequence 55, Appl
2	2388	100.0	8195	US-08-961-527-94	Sequence 94, Appl
3	987.6	41.3	2290	US-08-961-083-65	Sequence 65, Appl
4	653.8	27.4	2359	US-08-961-527-243	Sequence 243, Appl
5	385.4	16.1	1342	US-08-961-083-181	Sequence 181, Appl
6	385.4	16.1	6867	US-08-961-527-192	Sequence 192, Appl
7	381.4	16.0	973	US-08-961-527-355	Sequence 355, Appl
8	243.4	10.2	1684	US-08-961-527-258	Sequence 258, Appl
9	109	4.6	841	US-08-743-637B-34	Sequence 34, Appl
10	109	4.6	841	US-08-526-840B-34	Sequence 34, Appl
11	57.4	2.4	7218	US-08-232-463-14	Sequence 14, Appl
12	51.6	2.2	11091	US-08-134-001C-2243	Sequence 2243, Appl
13	45.2	1.9	2277	US-08-676-967-2	Sequence 2, Appl
14	45.2	1.9	2277	US-08-676-974-2	Sequence 2, Appl
15	45.2	1.9	2277	US-09-098-487-2	Sequence 2, Appl
16	41.4	1.7	5361	US-08-973-462-2	Sequence 2, Appl
17	41.4	1.7	6152	US-08-973-462-1	Sequence 1, Appl
18	41.2	1.7	696	US-08-461-687-193	Sequence 193, Appl
19	41.2	1.7	699	US-08-461-687-191	Sequence 191, Appl
20	41.2	1.7	717	US-09-461-697-189	Sequence 189, Appl
21	41.2	1.7	774	US-09-461-697-187	Sequence 187, Appl
22	41.2	1.7	819	US-09-461-697-185	Sequence 185, Appl
23	41.2	1.7	1669	US-09-461-697-184	Sequence 184, Appl
24	40.6	1.7	1704	US-08-134-001C-925	Sequence 925, Appl
25	40	1.7	30549	US-09-134-001C-322	Sequence 322, Appl
26	39.8	1.7	297	US-09-134-001C-1027	Sequence 1027, Appl
27	39.8	1.7	1716	US-09-134-001C-1028	Sequence 1028, Appl

28	39	1.6	1839	US-09-134-001C-1191	Sequence 1191, Appl
29	38.8	1.6	1856	US-08-360-606B-29	Sequence 29, Appl
30	38.8	1.6	2150	US-08-861-464-13	Sequence 13, Appl
31	38.8	1.6	2150	US-08-396-001-13	Sequence 13, Appl
32	38.8	1.6	2150	US-09-323-433A-13	Sequence 13, Appl
33	38.6	1.6	6744	US-08-119-125A-2	Sequence 2, Appl
34	38.6	1.6	8920	US-08-446-855A-1	Sequence 1, Appl
35	38.6	1.6	8920	US-09-150-741-1	Sequence 1, Appl
36	38.4	1.6	5319	US-08-169-927-1	Sequence 1, Appl
37	38.4	1.6	6846	US-08-961-527-198	Sequence 198, Appl
38	38	1.6	3280	US-08-259-000-4	Sequence 4, Appl
39	38	1.6	3280	US-08-729-767-6	Sequence 6, Appl
40	37.6	1.6	438	US-09-134-001C-1064	Sequence 1064, Appl
41	37.4	1.6	3294	US-08-923-992A-7	Sequence 7, Appl
42	37.4	1.6	26385	US-08-961-527-3	Sequence 3, Appl
43	37.2	1.6	1079	US-07-781-355-1	Sequence 1, Appl
44	37.2	1.6	2397	US-09-134-001C-2080	Sequence 2080, Appl
45	37	1.5	1103	US-09-122-400B-20	Sequence 20, Appl

## ALIGNMENTS

RESULT 1  
US-08-961-083-55  
Sequence 55, Application US/08961083  
Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESS: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2389 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-083-55

Query Match 100.0%; Score 2388; DB 3; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservativity 0; Mismatches 0; Indels 0;

0y 1 TTCTTACGAGTTGGGACTGTATCAAGCGTTAAGGAAATAATATCGTTTCTTA 60  
Db 1 TTCTTACGAGTTGGGACTGTATCAAGCGTTAAGGAAATAATATCGTTTCTTA 60

QY	61	TATGATGCGAAACAAACGACGCGAAAAACGGAGATTTGACTCCTCGATGAGGTATGCA	120
Db	61	TATGATGCGAAACAAACGACGCGAAAAACGGAGATTTGACTCCTCGATGAGGTATGCA	120
QY	121	GGGTGAAGGATTCATATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTAC	180
Db	121	GGGTGAAGGATTCATATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTAC	180
QY	181	TTTCACATGGCGACCACTATCATTTATACAAATGGTAAAGTTCCCTATAGACGTATCATG	240
Db	181	TTTCACATGGCGACCACTATCATTTATACAAATGGTAAAGTTCCCTATAGACGTATCATG	240
QY	241	TGAAGAATTACTCATGGAAGATCCAAACTTAAGCTAAAAAGATGAGATTTGTAATGA	300
Db	241	TGAAGAATTACTCATGGAAGATCCAAACTTAAGCTAAAAAGATGAGATTTGTAATGA	300
QY	301	GGTCAGAGGTGATATATTTATCAAGTAGATGGAATAATCTATGTAACTTAAGATATG	360
Db	301	GGTCAGAGGTGATATATTTATCAAGTAGATGGAATAATCTATGTAACTTAAGATATG	360
QY	361	TGCCACGCGGATTAAGCTCCGTACAAAAGAGAAATCAATGCACAAAACAAAGACATAG	420
Db	361	TGCCACGCGGATTAAGCTCCGTACAAAAGAGAAATCAATGCACAAAACAAAGACATAG	420
QY	421	TCACATGTGTAAAGTGGAACTCCAGAAACATGCTGCTGTGGCTTGGCACCTTGCCA	480
Db	421	TCACATGTGTAAAGTGGAACTCCAGAAACATGCTGCTGTGGCTTGGCACCTTGCCA	480
QY	481	AGGAGCTATATCTACAGATGATGGTATATCTTAAATGCTCGATATCATGAGATATC	540
Db	481	AGGAGCTATATCTACAGATGATGGTATATCTTAAATGCTCGATATCATGAGATATC	540
QY	541	TGTGTATGCTTATATCGTTCCCTCATGAGATCATTCACATTAATCTTAAGAAATGAGTT	600
Db	541	TGTGTATGCTTATATCGTTCCCTCATGAGATCATTCACATTAATCTTAAGAAATGAGTT	600
QY	601	ATCAGCTAGCAGATTTGGCTGCTGCAGAACGCTTCTATCTGTCGAGAAATCTGTCAA	660
Db	601	ATCAGCTAGCAGATTTGGCTGCTGCAGAACGCTTCTATCTGTCGAGAAATCTGTCAA	660
QY	661	TTTCAGAACCTATGCGCGACAAAATAGCGGTATACCTTCAAGAACAACTGGGTACCTTC	720
Db	661	TTTCAGAACCTATGCGCGACAAAATAGCGGTATACCTTCAAGAACAACTGGGTACCTTC	720
QY	721	TGTAAAGCATCCAGAACTCAAAATACTAACACAGCAACAGACAGACATCAACATGCA	780
Db	721	TGTAAAGCATCCAGAACTCAAAATACTAACACAGCAACAGACAGACATCAACATGCA	780
QY	781	AGCAAGTAAAGTATATACATTTGATAGTCTCTTGAACAGCTCTACAAATCTGCTTGA	840
Db	781	AGCAAGTAAAGTATATGATAGTCTCTTGAACAGCTCTACAAATCTGCTTGA	840
QY	841	TCACAGCATATTAATCTGATGGCCCTGTCTTGAATCAGACAAATACAAAGTCCGAC	900
Db	841	TCACAGCATATTAATCTGATGGCCCTGTCTTGAATCAGACAAATACAAAGTCCGAC	900
QY	901	AGCTAGAGTGTTCAGTGGCCACAGGAGATCATTAACCACTTCACTTACCTTACTCAAT	960
Db	901	AGCTAGAGTGTTCAGTGGCCACAGGAGATCATTAACCACTTCACTTACCTTACTCAAT	960
QY	961	GCTGTGAATTTGGAAGAACGATCGCTGTATTTATCCCTTCGTTATCGTTAAACCATTTG	1020
Db	961	GCTGTGAATTTGGAAGAACGATCGCTGTATTTATCCCTTCGTTATCGTTAAACCATTTG	1020
QY	1021	GGTACACGATTCAGGCGCAAGAACCAATCCACACCGCATCCGGAACTAGTCCAG	1080
Db	1021	GGTACACGATTCAGGCGCAAGAACCAATCCACACCGCATCCGGAACTAGTCCAG	1080
QY	1081	CCCGAATCTGACCAATATCTTAAATATAGCTTAAATTTCTTTGGTTAGTACCTGCT	1140
Db	1081	CCCGAATCTGACCAATATCTTAAATATAGCTTAAATTTCTTTGGTTAGTACCTGCT	1140
QY	1141	ACGAAAGTTGGGAGATATGATTTATCGAAGAAAGGCGCATCTCTCGTTATGCTTTGC	1200

Db	1141	ACGAAAGTTGGGGAGGATATGTATTCGAAAGAAAGGGCATCTCTCGTATGTCTTTGC	1200
Qy	1201	GAAGATTTTACCATCTGAAACGCTTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAG	1260
Db	1201	GAAGATTTTACCATCTGAAACGCTTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAG	1260
Qy	1261	TGTTTCACACACTTTAACTGCTTAAAAAGAAAATGTGCTCTCGTGACCAAGATTTTAA	1320
Db	1261	TGTTTCACACACTTTAACTGCTTAAAAAGAAAATGTGCTCTCGTGACCAAGATTTTAA	1320
Qy	1321	TGATAAGATATTAATCTGTCTAACGAGGCTCATTAAGCTCTGTTGNAAAATTAAGGGTCG	1380
Db	1321	TGATAAGATATTAATCTGTCTAACGAGGCTCATTAAGCTCTGTTGNAAAATTAAGGGTCG	1380
Qy	1381	TAAATCTGATTTTCCAAAGCCTTGTAGACAAATATTAGACGCTTGAATGATGATCACTTAA	1440
Db	1381	TAAATCTGATTTTCCAAAGCCTTGTAGACAAATATTAGACGCTTGAATGATGATCACTTAA	1440
Qy	1441	TAAAGAAAAATTGTAGATGATTTATTTGGCAATTCAGACCAATTTACCATCCAGACCG	1500
Db	1441	TAAAGAAAAATTGTAGATGATTTATTTGGCAATTCAGACCAATTTACCATCCAGACCG	1500
Qy	1501	ACCTGGCAACCAAAATCTCAATTAAGATATCTGAAGCGAAGTTCGATATGCTCAATT	1560
Db	1501	ACCTGGCAACCAAAATCTCAATTAAGATATCTGAAGCGAAGTTCGATATGCTCAATT	1560
Qy	1561	ACCTGTAAGTATACAACTGCAGATGGTTTATCTTTTGATGAACATGATATATAGTGA	1620
Db	1561	ACCTGTAAGTATACAACTGCAGATGGTTTATCTTTTGATGAACATGATATATAGTGA	1620
Qy	1621	TGAAGGAGATGATATGTATAGCCCTCATATGGCCATATGCTAGCTGATTTGAAAAAGATAG	1680
Db	1621	TGAAGGAGATGATATGTATAGCCCTCATATGGCCATATGCTAGCTGATTTGAAAAAGATAG	1680
Qy	1681	CCTTTCTGTATAGGAAAAAGTTGCGAGCTCAACCCCTATACTTAAAGAAAAAGATATCTTACC	1740
Db	1681	CCTTTCTGTATAGGAAAAAGTTGCGAGCTCAACCCCTATACTTAAAGAAAAAGATATCTTACC	1740
Qy	1741	TCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATATGTGACGACGATTTTAA	1800
Db	1741	TCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATATGTGACGACGATTTTAA	1800
Qy	1801	CAATCGTGTGAAAAGGGGAAAAAGATTCCTCACTGTTGCATCTTCATATATGTGTTGAGCA	1860
Db	1801	CAATCGTGTGAAAAGGGGAAAAAGATTCCTCACTGTTGCATCTTCATATATGTGTTGAGCA	1860
Qy	1861	TACAGTTGAGGTTAAAAACGGTAAATTTGATTTCTCATTAAGGATTCATTTACATAAATAT	1920
Db	1861	TACAGTTGAGGTTAAAAACGGTAAATTTGATTTCTCATTAAGGATTCATTTACATAAATAT	1920
Qy	1921	TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGACTTACCTTGGGAAGA	1980
Db	1921	TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGACTTACCTTGGGAAGA	1980
Qy	1981	TTTGTTTTGGCAGATTAAGTACTAGCTAGTAACACCCTGACGAACGCTCCACATTTCAATGA	2040
Db	1981	TTTGTTTTGGCAGATTAAGTACTAGCTAGTAACACCCTGACGAACGCTCCACATTTCAATGA	2040
Qy	2041	TGGATTTGGGCAATGCCAGTGAACATGTGTTAGGCAAGAAAGCACCACTGAAGATCCAAA	2100
Db	2041	TGGATTTGGGCAATGCCAGTGAACATGTGTTAGGCAAGAAAGCACCACTGAAGATCCAAA	2100
Qy	2101	TAAAGAACTTCAAAAGGGGATGAAGACGACGTAGAGAAACACTGCTGTAGCCAGAAAGTCC	2160
Db	2101	TAAAGAACTTCAAAAGGGGATGAAGACGACGTAGAGAAACACTGCTGTAGCCAGAAAGTCC	2160
Qy	2161	TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAAAGACGAAGATTTTGCTTGGCAA	2220
Db	2161	TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAAAGACGAAGATTTTGCTTGGCAA	2220
Qy	2221	AGTAAACGATTTCTAGTCTGAAAAGCCAAATGCAACAGAAACTTACGTGGTTTACGAAATTA	2280
Db	2221	AGTAAACGATTTCTAGTCTGAAAAGCCAAATGCAACAGAAACTTACGTGGTTTACGAAATTA	2280

Db 2221 AGTAAGGATTCGTAGTGTGAAGCCAAATGCAACGAAGTCTAGCTGTTAGCAAAATA 2280  
QY 2281 TTTCACCTCAATATTTGATAGATAGTATGATGCGAGAGCAAAAAATTTACTTGC 2340  
Db 2281 TTTCACCTCAATATTTGATAGATAGTATGATGCGAGAGCAAAAAATTTACTTGC 2340  
QY 2341 GTTGTAAAGGAAGTAACTCTCATCTGTAGTAGGAAAAATAAAC 2389  
Db 2341 GTTGTAAAGGAAGTAACTCTCATCTGTAGTAGGAAAAATAAAC 2389

## RESULT 2

US-08-961-527-94  
; Sequence 94, Application US/08961527  
; Patent No. 6420135  
GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8195 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-94

Query Match 100.0%; Score 2388; DB 4; Length 8195;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGACGGTTAAAGAAAAATATCGTGTTCCTTA 60  
Db 3053 TTCTTACGAGTTGGAGCTGTATCAAGCTAGACGGTTAAAGAAAAATATCGTGTTCCTTA 3112  
QY 61 TATAGATGGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 120  
Db 3113 TATAGATGGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3172  
QY 121 GCGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGCAAGCAAGCAAGCAAG 180  
Db 3173 GCGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGCAAGCAAGCAAGCAAG 3232  
QY 181 TTTCATGCGGACACATATATTTTACAAATGTTAGGTTCTTATGAGCCTATCATCAG 240  
Db 3233 TTTCATGCGGACACATATATTTTACAAATGTTAGGTTCTTATGAGCCTATCATCAG 3292

QY 241 TGAAATTTACTCTGAAGAGTCCAAACTATATAGCTTAAAGATGAGGATATTTTATGA 300  
Db 3293 TGAAATTTACTCTGAAGAGTCCAAACTATATAGCTTAAAGATGAGGATATTTTATGA 3332  
QY 301 GGTCGAAGGATGATATGTTATCAAGGATAGATGAAAAATCTATGTTTAAAGATGCG 360  
Db 3353 GGTCGAAGGATGATATGTTATCAAGGATAGATGAAAAATCTATGTTTAAAGATGCG 3412  
QY 361 TGCCACGCGGATATAGCTCGGTACAAAAAGGAATCAATGCACAAAAACAAGCATAG 420  
Db 3413 TGCCACGCGGATATAGCTCGGTACAAAAAGGAATCAATGCACAAAAACAAGCATAG 3472  
QY 421 TCAACATCGTGAAGGATGGAAGTCCAAAGCAAGATGATGCTGCTTGGCAGCTTGCA 480  
Db 3473 TCAACATCGTGAAGGATGGAAGTCCAAAGCAAGATGATGCTGCTTGGCAGCTTGCA 3532  
QY 481 AGGAGCTATATCTACAGATGATGTTATCTTAAATGCTTGTATCATATAGAGATAC 540  
Db 3533 AGGAGCTATATCTACAGATGATGTTATCTTAAATGCTTGTATCATATAGAGATAC 3592  
QY 541 TGGTATGCTTATATGCTTCTCATGAGATCATACCATTTACCTCTTAAGATGAGTT 600  
Db 3593 TGGTATGCTTATATGCTTCTCATGAGATCATACCATTTACCTCTTAAGATGAGTT 3652  
QY 601 ATCAGTAGGAGTGGCTGCTGCGAAGCCTTCTATCTGCTGAGAGAAATCTGTCAA 660  
Db 3653 ATCAGTAGGAGTGGCTGCTGCGAAGCCTTCTATCTGCTGAGAGAAATCTGTCAA 3712  
QY 661 TTCAAGAACCTATTCGCGCAAAAAATATAGCATATACATCTCAAGAACAGTGGATCTTC 720  
Db 3713 TTCAAGAACCTATTCGCGCAAAAAATATAGCATATACATCTCAAGAACAGTGGATCTTC 3772  
QY 721 TGTAGCAATCCAGGAATCAAAATACTAAACACAGCAACAGCAACACTAACAGTCA 780  
Db 3773 TGTAGCAATCCAGGAATCAAAATACTAAACACAGCAACAGCAACACTAACAGTCA 3832  
QY 781 AGCAAGTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
Db 3833 AGCAAGTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3892  
QY 841 TCAAGCATGTAGAAATCTGATGCGCTTGTCTTGTATCCAGCACAAATCAACAGTCGAC 900  
Db 3893 TCAAGCATGTAGAAATCTGATGCGCTTGTCTTGTATCCAGCACAAATCAACAGTCGAC 3952  
QY 901 AGCTAGAGGTTGTCAGATGCCACAGGAGATCATACCATCTCATCTCTCAAT 960  
Db 3953 AGCTAGAGGTTGTCAGATGCCACAGGAGATCATACCATCTCATCTCTCAAT 4012  
QY 961 GTCGAAATGGAAGAAAGAAATGCTGATTTATTCGCTCGTTATGCTCAACCAATG 1020  
Db 4013 GTCGAAATGGAAGAAAGAAATGCTGATTTATTCGCTCGTTATGCTCAACCAATG 4072  
QY 1021 GGTACGATTTCAAGGCGCAGACAAACCAAGTCGACACGACATCCGAGACTAGTCCAG 1080  
Db 4073 GGTACGATTTCAAGGCGCAGACAAACCAAGTCGACACGACATCCGAGACTAGTCCAG 4132  
QY 1081 CCGCAACCTGACCAAAATCTTAAATATGACTCAAAATCTTGTGTAGTCACTGCT 1140  
Db 4133 CCGCAACCTGACCAAAATCTTAAATATGACTCAAAATCTTGTGTAGTCACTGCT 4192  
QY 1141 ACGAAAGTGGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
Db 4193 ACGAAAGTGGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 4252  
QY 1201 GAAAGATTTACATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260  
Db 4253 GAAAGATTTACATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 4312  
QY 1261 TGTTCACACACTTTAACTGCTTAAAAAAGAAATGTTGCTCTGCTGACCAAGATTTTA 1320  
Db 4313 TGTTCACACACTTTAACTGCTTAAAAAAGAAATGTTGCTCTGCTGACCAAGATTTTA 4372  
QY 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTGMAATTAAGGCTG 1380

Db	4373	TGATTAAGCATATATCTGTTAACTGAGCTCATTAAGCCTTGTTGAAATAAGGCTGC	44322
Qy	1381	TAATTTGATTTCCAGCCTTTAGACAATAATTTAGAACGCTTGAATGATGATTCGACTTA	14400
Db	4433	TAATTCGATTTCCAAACCTTTAGACAAATTTATAGAACCTTGAAATGATGATGACTTA	44922
Qy	1441	TAAAGAAAAATTGGTAGATATTTATTTGGCATTTCCTAGACCAATTTACCATCCAGACG	15000
Qy	1501	ACTTGCGAAACCAAAATTCCTAAATTTAGATATCTGTAAGACGAAGTTCGATGGCTCAAT	15600
Db	4493	TAAAGAAAAATTTGGTAGATATTTATTTGGCATTTCCTAGACCAAAATTTACCATCCAGACG	44522
Qy	1501	ACTTGCGAAACCAAAATTCCTAAATTTAGATATCTGTAAGACGAAGTTCGATGGCTCAAT	15600
Db	4553	ACTTGCGAAACCAAAATTCCTAAATTTAGATATCTGTAAGACGAAGTTCGATGGCTCAAT	46122
Qy	1561	AGCTGATPAGTATACAGCGTCAGATGGTTACATTTTGTGTAACATGATATATACATGA	16200
Db	4613	AGCTGATPAGTATACAGCGTCAGATGGTTACATTTTGTGTAACATGATATATACATGA	46722
Db	4673	TGAAGGAGATGCATATGTAAAGCCTCATATGGGCCATATCTACTGGATTTGAAAGATAG	47322
Qy	1681	CCTTTCGATAGAGAAAAATTTGCAGCTCAAGCCATATAGAAAAAAGATATCTTACC	17400
Db	4733	CCTTTCGATAGAGAAAAATTTGCAGCTCAAGCCATATAGAAAAAAGATATCTTACC	47922
Qy	1741	TCGATCTCCAGACGCAGATGTTAAAGCAATCCAACTGAGATGATGCGACGCTATTTA	18000
Db	4793	TCGATCTCCAGACGCAGATGTTAAAGCAATCCAACTGAGATGATGCGACGCTATTTA	48522
Qy	1801	CAATCGTGTGAAGGGGAAAAAGCAATTCACGCTGTCAGCTCCATATATATGAGTTGACCA	18600
Db	4853	CAATCGTGTGAAGGGGAAAAAGCAATTCACGCTGTCAGCTCCATATATATGAGTTGACCA	49122
Qy	1861	TACAGTTGAGGTTAAAAACCGTAAATTTGATATTTCCCTAAAGATCATTTACCATATAT	19200
Db	4913	TACAGTTGAGGTTAAAAACCGTAAATTTGATATTTCCCTAAAGATCATTTACCATATAT	49722
Qy	1921	TAAATTTGCTTGGTTTGATATATACACATATCAAAAGCTCCAAATGGCTATATACCTTGGAGA	19800
Db	4973	TAAATTTGCTTGGTTTGATATATACACATATCAAAAGCTCCAAATGGCTATATACCTTGGAGA	50322
Qy	1981	TTTGTTTGGCGAGTTTATAGTACTAGTATAGAACACCCCTGAGCAAGCTCCACATTTCTAATGA	20400
Db	5033	TTTGTTTGGCGAGTTTATAGTACTAGTATAGAACACCCCTGAGCAAGCTCCACATTTCTAATGA	50922
Qy	2041	TGATGAGGGGCAATGCCAGTAGCATGTGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA	21000
Db	5093	TGATGAGGGGCAATGCCAGTAGCATGTGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA	51522
Qy	2101	TAAAGACTTCAAAAGCGGATGATAGAGCCAGTATAGAGAAACCTGCTGAGCCAGAAGTCC	21600
Db	5153	TAAAGACTTCAAAAGCGGATGATAGAGCCAGTATAGAGAAACCTGCTGAGCCAGAAGTCC	52122
Qy	2161	TCAAGTATGAGAGCTGAAAAAATATAGAACCCACATCAAAAGAGAGAAAGTTTGGCTTGGCAA	22200
Db	5213	TCAAGTATGAGAGCTGAAAAAATATAGAACCCACATCAAAAGAGAGAAAGTTTGGCTTGGCAA	52722
Qy	2221	AGTAAAGCATTTAGTCTGTAAGAACCCCAATCAACAGAAACTAGCTGGTTTACGAAATAA	22800
Db	5273	AGTAAAGCATTTAGTCTGTAAGAACCCCAATCAACAGAAACTAGCTGGTTTACGAAATAA	53322
Qy	2281	TTTGTACTCTTCAAAATTTATGGATATACAAATATCATATCATGCGAGAGACAGAAAAATTAATTCTTGC	23400
Db	5333	TTTGTACTCTTCAAAATTTATGGATATACAAATATCATATCATGCGAGAGACAGAAAAATTAATTCTTGC	53922
Qy	2341	GTTGGTTAAAGGAAGTATACCTTCAATCTTATAGTAAAGAAAAAATAAAC	2389
Db	5393	GTTGGTTAAAGGAAGTATACCTTCAATCTTATAGTAAAGAAAAAATAAAC	5441

```

US-08-961-083-65
; Sequence 65, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-65

```

Query Match	41.3%	Score 987.6	DB:3	Length 2290
Best Local Similarity	67.7%	Pred. M: 9,6e-262		
Matches 1481	Conservative	0	Mismatches 645	Indels 60
				Gaps
QY	1	TTCTTCAGAGTTGGACCTGTATCAAGCTTAGACGGTTTAAGGAAAA--TAATCGTGTTC	57	
Db	4	TTCCATATGACTTGGTCGTCAACCAAGCTGTAGGTTAAGAAAAGCTAATCGAGTTTC	63	
OY	58	CTATATAGATGTAANAACAGCAGCCAAAAAAGGAAATTTGACTCTGATGAGGTG	117	
Db	64	TTATATAGATGGTGTATCGCTGCACAAAAGGCAAAAACCTTGACACCGAATGAAGTCA	123	
OY	118	CAAGGGTGAAGATCAATATGCTGAGCAATATCTCAATCAATACAGACCAAGCATATG	177	
Db	124	TAAAGGGAGGGGATCAACGCCGCAACAAATNTNTATCAAGATTTACGGATCAAGGTTATG	183	
OY	178	CACTTACATGGCGACCACTATCATTTATTAACATGTAAGGTTCCTTATGACGCTATCAT	237	
Db	184	GACCTCATATGGAGACCTTATCATTTACTATATGGAAGTTCCTTATGATGCGATCAT	243	
OY	238	CAGTAAAGATTTACTCATGAAGAAATCCAAACTATATAGCTAAAGAAAGTAAAGGATATGTTAA	297	
Db	244	CAGTAAAGAGCTCCATCAATGAAGAAATCCGAATTTATCAAGTGAAGGATTCAGAAATGTCTCA	303	
OY	298	TGAGGTCAAGGGTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACTTTAAGGA	357	
Db	304	TGAATTCAAAGGTGGTTATGTCATTAAAGGTAACGGTAAATCTAATGTTTACTTTAAGGA	363	
OY	358	TGCTCCCAACGCGGATTAACGTCCTGATCAAAAGAGGAATCAATCGACAAAAACAAGACA	417	
Db	364	TGCACCTCATGGGTAATATATTCGACAAAGAAAGATTAATCAACCTTCAGACACAGAACG	423	

OY	418	TAGTCACATGCTGAAGGTGGAGCTCCAGAAACAGAGTGGTGGTGGTGGTGGCAGCTTC	477
Db	424	CAGTCATATATCAT-----AAGTCAGAGCAGATTAATGCTGTGGCAGCCAGAGC	474
OY	478	GCAAGAGCGCTATCTACAGATGATGGTTATATCTTTAACTGGCTTGATATCATGAGAGA	537
Db	475	CCAAAGAGCTTATATCAACAGGATGATGGTATATCTTCATATGATCTGATATCATATGAGAGA	534
OY	538	TACTGTGATGCTTATATATGCTTCCATGGAGATCATTCACATTCATTCCTAAGAGATGA	597
Db	535	CACGGGTGATGCTTATATTCGTTCTTCACAGGCGACATTCACATTCATTCCTAAGAGATGA	594
OY	598	GTTTATCAGCTAAGCGAGTGGCTGCTGCAAGAGCTTCCATCTGCTGAGAGAAATCTGTG	657
Db	595	GTTTATCAGCTAAGCGAGTGGCTGCTGCAAGAGCT-----	629
OY	658	AAATTCAGAACTATGCGCGAGCAAAATATAGCCATTCGTAACAGCAAACTGGGTACC	717
Db	630	-----ATTGGAATGGGAGAGCAGGAGATCTGCTCTTTTAACTTCTAGTTAATATGC	681
OY	718	TTTCTAAGCAATCCAGAACTCAAAATATCAACAGCAAAACAGACACATCAACAG	777
Db	682	AAATTCAGCTCAACCAAGATTTGTCAGAGAACCAACATCTGATGTCATCCACATTA	741
OY	778	TCAAGCAAGTCAAAAGTAAATACATCTGATAGTCTCTTGAAGACGCTCTACAAATCCCTT	837
Db	742	TCA-----AAATTCAGGGGAGAAACATTCACAGCCCTTTACGTGAATTTAGCTAAACCTT	798
OY	838	GAGTCACAGCATATGATAGAAATCTGATGGCTCTGTCTTGTATCCAGCAAAATACAGAGC	897
Db	799	ATCAGAAAGCGCATATGGAATCTGATGCGCTTATTTTCGACCAGGCGCAATTCACAGAGC	858
OY	898	AACAGCTAAGAGTGTTCAGTGTCCACAGGAGATCATTCACACTTCATCCTTACTCTCA	957
Db	859	AAACCGCAGAGGTGTAGCTGTCCCTCATGTTAGTAAACATTCACACTTATCCCTTATGAGACA	918
OY	958	AATCTCTGAATTTGGAGAAAGCAAGATTCCTGCTATATATCCCTCTGTTATCGTTCAACCA	1017
Db	919	AATGCTGTGAATTTGGAGAAAGCAATTTCTCTGTATATATCCCTCTGTTATCGTTCAACCA	978
OY	1018	TTTGGGTACCAAGATTCAGAGCCCAAGAACACCAAGTCCACAGCAGCTCGGAAACCTAGTCC	1077
Db	979	TTTGGGTACCAAGATTCAGAGCCCAAGAACACCAAGTCCACATTCGATCCGGAACCTAGTCC	1038
OY	1078	AGGCGCGCAACCTGACACCAATCTTAAATATAGCTCAAAATCTTC-----TTT	1125
Db	1039	AAGTCGGAACCTGACACCAATCTCTCAACAGCTCCACACCAATTCATATGAGAAAT	1098
OY	1126	GGTATGTCAGCTGTTACGAAAGTTGGGGAGAGATATGTTATTCGAGAGAAAGGCGATCTG	1185
Db	1099	GGTCAAGAAAGCTGTTTCGAAAGTAGGCGCATGGTTATGCTTTGAGAGAAATGAGGTTTC	1158
OY	1186	TGCTGATGCTTTGCGAAGATTTTACCATCTGAAATCTGTTAAATCTTGAAGCAAGTT	1245
Db	1159	TGCTGATATTCACAGCCCAAGGATCTTTCAGAGAAACAGAGCAGGCTATGATAGCAAACT	1218
OY	1246	ATCAAAACAAGAGAGTGTTCACACACTTTAACTGTAATAAAGAGAAATGTTGCTCCG	1305
Db	1219	GGCCACAGAGAAAGTTTATCTATCATAGCTAGSAGCTAAGAAACTGACCTCCCATCTAG	1278
OY	1306	TGACCAAGAAATTTTATGATATACCATATATATCTGTTAACTGAGGCTCATTAAGCCCTGTT	1365
Db	1279	TGATCGAGAAATTTTACAAATTAAGGCTTATACCTTACTAGCAAGAAATTCACCAAGATTACT	1338
OY	1366	TGNAATATTAAGGCTGCTAATTCGATTTCCAACTTACAGCCTTAGCAAAATATTTAGAACGTTGAA	1425
Db	1339	TGATATATTAAGGCTGACAGAGTATTTTGAAGCTTTGGATTAACCTGTTGGAACGCTCAA	1398
OY	1426	TGATGATGCACTAATTAATAAGAAATTTGGATAGCTTATATGGCATCTTACAGCAAT	1485
Db	1399	GGATGTGTCAAAGTATTAATCAAGTTAGTGGAGNATATCTTGCCTCTTATAGCTCCGAT	1458
OY	1486	TACCATTCACAGAGCTTGGCAACCAATTCATAATTTGAGTATCTAGAGAGCAAGT	1545

[illegible]



```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36, 373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 181:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-083-181

Query Match      16.1%; Score 385.4; DB 3; Length 1342;
Best Local Similarity 65.3%; Pred. No. 2.8e-96;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

37 TAAGGAAATATCGTTCTCTATATAGATGGAACAAGCGCAAAAAGAGAA 96
|||||
28 TAAGGACAAATATCGTCTCTATATGTGACAGCGACAGTCAAGTCAGAAAGTGA 87
|||||
97 TTGACTCGATGAGGTAGCAAGCGGTGAAGCAATCATGTCAGCAAAATCGTCAC 156
|||||
88 CTGACACACGACGAGGTAGCCAGAAAGAAATTCAGGCTGACCAATTTGATCAA 147
|||||
157 GATACAGACCAAGGCTATGCTACTACATGCGACCACTATCATTTATACATGTA 216
|||||
148 AATTAGAGATCAGGCTATGCTATAGCTACACGCTGACCACTATCATTTATGGA 207
|||||
217 GGTCTTATGACGCTATCATGAGTGAAGATTTACTCATGAAAGATCCAACTATA 276
|||||
208 AGTTCTTATGATGCGCTCTTAGTAGAAGAACTTGATGAGAGATCCAACTATCA 267
|||||
277 AAAAGATGAGATATGTTATATAGGTCAAGGTGATATGTTATCAAGTATGGA 336
|||||
268 TAAAGACGCTGATATGTCATGAAGTCAAGGCTGTTATATCATCAAGTCAATG 327
|||||
337 ATACTATGTTTAAAGATGCTGCCACGCGGATTAAGCTCCGTAACAAAGAGAA 396
|||||
328 ATATATGCTTACTGTAAGATGACGCTCATGCTGATTAATGTTGCAATTAAGAT 387
|||||
397 CAATGACAAAACAAGACATAGTCAACATCGTGAAGGTGAATCCAGAAACGATG 456
|||||
388 CAATCGTCAAAAACAAGATGTCAAAGATATGAAGGTTA-----ACTC 435
|||||
457 TCGTGTGCTTGGCAGCTGCGAAGAGCGCTTACTACAGATGATGTTATCTTTAA 516
|||||
436 TAAATGCTGTAAGCAAGGCTCAGGAGCATTAACGAAAGATGATGTTATCTTT 495
|||||
517 TCGTGTGATATCATAGAGATGCTGATGCTTATATGCTTCTCATGAGATATTA 576
|||||
496 TCCAGCTGATATATGAGATACGGGTAATGCTTATATGCTTCTCATGAGAGTAC 555
|||||
577 CCATTACATTCCTAAGATGAGTTATACCTAGCGAGTTGGCTGTCAGAAAGCTT 636
|||||
556 TCACCTACATTCCTCAAAAGGATTTATCTGATGTAATTAGCAGAGCTTAACAC 615
|||||
637 ATCTGTCAGAGAAATCTGCAATTCAGAACTTCCGCCGCAAAATGCGATTAAC 696
|||||
616 GGTGGAATAAATATGCAACCGAGTCAAGTTAAGCTATTTCTTCAACAGCTAGT 672
|||||
697 TTCAAGAACAACTGGGTACTTCTGTAAGCAATCCAGAACTACAAATATTAACA 756
|||||
673 -----TAAACGCAATCTGT 687
|||||
757 CAACAACAGCACTTAACAGTCAAGCAAGTCAAAATATGACATTTGATAGTCTG 816
|||||
688 AGCAAAAGATCAACTAGCAAGCAAGCAAAATTAATCTGAAATATCTCCAGAGT 747
|||||
```

```

QY 817 ACAGCTACAAACTGCTTTGAGTCAAGCATAGTAAATCTGATGCGCTGCTTGA 876
| | | | |
DB 748 GGAAGCTATGATTCACCTAGCCGCCACAGCTTACAGTGAATCGATGCGCTTGA 807
| | | | |
QY 877 TCCAGCAAAATCAACAGTCAAGTCAAGTGTGCTGAGTCCACAGCAGATCATTA 936
| | | | |
DB 808 CCTGCTAATATATCATGCTGATACCAAAATGAGAGTTGCGATTCGAGATGCCATTA 867
| | | | |
QY 937 CCACTTCATCCCTTACTTCAAAATGCTGAATTTGAAGAAAGAAATGCTGATTA 996
| | | | |
DB 868 CCACTTATTCCTTACAGCAAGCTTCTGCTTGAAGAAAGAAATGCTGAGATG 927
| | | | |
QY 997 CCTTGT 1003
| | |
DB 928 TATCAGT 934

RESULT 6
US-08-961-527-192/c
; Sequence 192, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36, 373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-192

Query Match      16.1%; Score 385.4; DB 4; Length 6867;
Best Local Similarity 65.3%; Pred. No. 6.5e-96;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATATCGTTCTCTATATAGATGGAACAAGCGCAAAAAGAGAA 96
|||||
DB 6887 TAAGGACAAATATCGTCTCTATATGATGACAGCAAGTCAAGTCAAGTGA 6628
|||||
QY 97 TTGACTCTGATGAGCTTAGCAAGCGTGAAGAAATCATATGCTGACAAATCTCT 156
|||||
DB 6627 CTGACACACGACGAGCTTAGCCAGAAAGAAATTCAGGCTGACCAATTTG 6568
|||||
QY 157 GATACAGACCAAGGCTATGCTACTTCAATGCGACCACTATCATTTATACATG 216
|||||
```





QY 598 GTTATCAGCTAGCGAGTGGCTGCTGCAGAGCCT 632  
|||||  
Db 748 GTTATCAGCTAGCGAGTGGCTGCTGCAGAGCCT 782

## RESULT 8

US-08-961-527-258  
; Sequence 258, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunach  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 258:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1684 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-258

Query Match 10.2%; Score 243.4; DB 4; Length 1684;  
Best Local Similarity 60.3%; Pred. No. 3.5e-57;  
Matches 493; Conservative 0; Mismatches 236; Indels 88; Gaps 2;

QY 1629 ATGCATATGTAAGCGCTCATAGGCGCATAGTACAGTGGATTTGAAAAAGATAGCCTTTCG 1688  
|||||  
Db 1 ATGCCATATGTAAGCGCTCATAGGCGCATAGTACAGTGGATTTGAAAAAGATAGTTCG 60  
QY 1689 ATAAAGAAAAAGTTGCGAGTCAAGCCTATACATACTAAAGAAAAAGTATCCATCCATCTC 1748  
|||||  
Db 61 AACCTAGAGAGAGGGCA-CCAGGCTTATGCTAAAGAGAGTGGACCCCTCTTCA 119  
QY 1749 CAGACGCAATGTTAAAGCAATCAATGAGATAGTGCAGCAGCATTTTACCAATCGTG 1808  
|||||  
Db 120 CAGACCATGAGATTCAGAAATACAGGCAAAAGAGAGAGCAAGATCTACCAACCGG 179  
QY 1809 TGAAGGGAAAAAGCAATTCACGCTGTCGATTCATATATAGTTGAGCATACAGTTG 1868  
|||||  
Db 180 TGAAGCGACTAAGAGGTGCGCACTTGATGTAATGCTTCAATTCATATATAGTTAG 239  
QY 1869 AGCTTAAAAAGCTAATTTGATTTCTCATTAAGATCATTTACATTAATTAATTTG 1928  
|||||  
Db 240 AACTCAAAAAAGCTAATTTGATTTCTCATTAAGATCATTTACATTAATTAATTTG 299  
QY 1929 CTTGGTTGATGATCAACATACAAAGCTCCAAATGCTATACCTTGGAAGATTTGTTTG 1988  
|||||

Db 300 AGTGGTTTGACGAAGCGCTTTATAGGACACCTAAGGGTATACCTTGAGGATCTTTTG 359  
QY 1989 CGACGATTAAGTACTAGTGAACACCCCTGACGACGCTCCATTTCTAATGATGATGG 2048  
|||||  
Db 360 CGACTGTCAAGTACTATGTGTGAACATCCAAACGACGCTCCGCTTCAAGTATATGTTTG 419  
QY 2049 GCAATGCCAGTGAAGATGTGTTAGGCAAGAAAGACCACTGAGATCCAAATGAAGACT 2108  
|||||  
Db 420 GTAACGCTAGCGACATGTTCAAAGAAACAAATGGTCAAGCTATACCAATCAAAAG 479  
QY 2109 TCAAGCGATGA----- 2121  
Db 480 AAAAACCAGCGAGAGAAACCTCAGACAGAAAAACCTGAGAAAGAAACCTCGAAG 539  
QY 2122 -----AGACGCAGTGAGGAAACAC 2141  
Db 540 AGAACCAGCAAGCGAGAAACGAGTCTCCAAACCAGAGAGACCGAAGATAC 599  
QY 2142 CTGCTAGCGCAGAGTCCCTCAAGTAGAGACTGAAAGAGGCCCAACCAAGAG 2201  
Db 600 CAGAGGATCAGAAAGACCTCAGAGTGAAGAGTTGAAGAAACTGAGAGAG 659  
QY 2202 CAGAGTTTCTGCGGAAAGTACGAGATTTAGTCTGAAAGCCAAATGCAACGAAACTC 2261  
|||||  
Db 660 CTGAAAGATTTACTTGGAAAAATCCAGGATCCAAATTAAGTCAAGTCCAAATGCCAAAGAGACTC 719  
QY 2262 TACCTGTTTACGAATTAATTTGCTCTCAATTTATGATTAACATTAATGAGCAG 2321  
Db 720 TCACAGGATTAATAAATTAATTTGCTGACCCAGGACCAAACTATTAATGAGCAG 779  
QY 2322 AAGCAGAAAAATTAATTTGCTGTTTAAAGAGAGTAA 2358  
Db 780 AAGCTGAAAAATTAATTTGCTTATTAAGAGAGTAA 816

## RESULT 9

US-08-743-637B-34/C  
; Sequence 34, Application US/08743637B  
; Patent No. 5994066  
; GENERAL INFORMATION:  
; APPLICANT: BERGERON, Michel G.  
; APPLICANT: PICARD, Francois J.  
; APPLICANT: OUELLETTE, Marc  
; APPLICANT: ROY, Paul H.  
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
; IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
; NUMBER OF SEQUENCES: 273  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: QUARLES & BRADY  
; STREET: 411 EAST WISCONSIN AVENUE  
; CITY: MILWAUKEE  
; STATE: WISCONSIN  
; COUNTRY: USA  
; ZIP: 53202-4497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/743,637B  
; FILING DATE: 04-Nov-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/526,840  
; FILING DATE: 11-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BAKER, Jean C.  
; REGISTRATION NUMBER: 35,433  
; REFERENCE/DOCKET NUMBER: 850586.90012  
; TELECOMMUNICATION INFORMATION:

: TELEPHONE: (414) 277-5000  
 : TELEFAX: (414)277-5591  
 : INFORMATION FOR SEQ ID NO: 34:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 841 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: DNA (genomic)  
 : ORIGINAL SOURCE: Streptococcus pneumoniae  
 : OS-08-743-637B-34

Query Match	4.6%	Score 109	DB 2	Length 841
Best Local Similarity	65.4%	Pred. No. 2.2e 20		
Matches 176	Conservative	0	Mismatches 90	Indels 3
				Gaps 1

RESULT 10  
 US-08-526-840B-34/c  
 Sequence 34, Application US/08526840B  
 Patent No. 6001564  
 GENERAL INFORMATION:  
 APPLICANT: BERGERON, Michel G.  
 APPLICANT: OUELLETTE, Marc  
 APPLICANT: ROY, Paul H.  
 TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND  
 TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY  
 TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE  
 TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...  
 NUMBER OF SEQUENCES: 177  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OUELLETTE & BRADY  
 STREET: 411 East Wisconsin Avenue  
 CITY: Milwaukee  
 STATE: Wisconsin  
 COUNTRY: USA  
 ZIP: 53202-4497  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/526,840B  
 FILING DATE: 11-SEP-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/304,732  
 FILING DATE: 12-SEP-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BAKER, Jean C.  
 REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 850586.90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 841 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: *Streptococcus pneumoniae*  
J5-08-526-840B-34

Query Match	4.68;	Score 109;	DB 3;	Length 841;
Best Local Similarity	65.48;	Pred. No. 2.2e-20;		
Matches 176; Conservative	0;	Mismatches 90;	Indels 3;	Gaps 1

RESULT 11  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZgpt-F1s  
US-08-232-463-14

Query Match 2.4%; Score 57.4; DB 1; Length 7218;  
Best Local Similarity 3.4%; Pred. No. 1e-05;  
Matches 13; Conservative 224; Mismatches 150; Indels 0; Gaps 0;

QY 1974 TGGAGATTGTTTGGCATTAACTAGTACAGACACCCCTGACGAGCAGTTCACATT 2033  
| ||||| | | | :  
DB 1450 TAGAACAATTGTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391  
QY 2034 CTAAATGATGATGGGCAATGCCAGTACAGTGTGTAGCAGAAAGACCAAGTGAAG 2093  
:  
DB 1390 RRR 1331  
QY 2094 ATCCAAATAGAACTTCAAGGGATGAAGAGCAGCTAGAGAAACACTGCTGAGCCAG 2153  
:  
DB 1330 RRR 1271  
QY 2154 AAGTCCCTCAAGTAGAGTGAAGAGCCCACTCAAGAGAGAGAGAGTTCGC 2213  
:  
DB 1270 RRR 1211  
QY 2214 TTGCGAAGTAAAGATTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGTTTAC 2273  
:  
DB 1210 RRR 1151  
QY 2274 GAATATATTGACTCTTCAATTAATGATACAAATAGTTCATGCGAGAGAGAGAAAT 2333  
:  
DB 1150 RRR 1091  
QY 2334 TACTGCGTTGTTAAAGAGAGTAATC 2360  
:  
DB 1090 RRR 1064

RESULT 12  
US-09-134-001C-2243  
Sequence 2243, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2243  
LENGTH: 11091  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2243

Query Match 2.2%; Score 51.6; DB 4; Length 11091;  
Best Local Similarity 47.3%; Pred. No. 0.0005;

Matches 187; Conservative 0; Mismatches 205; Indels 3; Gaps 1;

QY 1096 AATCTTAAATAGACTCAATCTCTTGGTTACTAGCTGTACGAAAGTTGGGA 1155  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8481 AAAAGCAAAATGATGTAATCAATCAACTAATCAAGCAAGTGAATCTGAGCA 8540  
QY 1156 AGGATATGATTCGAGAAAGGAGCATCTCTGTTATGCTTTGCGAAA---GATTACC 1212  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8541 AATATTTAGATCAATCAATCAATCACTCAAGACAGATTTTAAAGAAAGTATGACAG 8600  
QY 1213 ATCTGAAGTGTAAATATCTGGAAGCAAGTATCAAAAACAGAGAGTGTTCACACAC 1272  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8601 AGCTGAATTTGTTAAAGCCCAACAAACAAATTTGATGAATAGACAGAAATTTAGTGC 8660  
QY 1273 TTTAACTGTAAAGAAAGAAATGCTCTCTGTCGACCAAGATTTTATGATTAACATA 1332  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8661 TACACAGAGGAAAGAAAGACATGCTTTACAACTTTAGATGAACAGCTTAACAAATTCAT 8720  
QY 1333 TATCTGTTAACTGAGGCTCATTAAGCCTGTTGNAATTAAGGCTGTAATTCGATTT 1392  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8721 TATTTATTAATCAAGCTAATACATATATAGATGATTAATGCTAAACCTTCTGGTT 8780  
QY 1393 CCAAGCCTTAGACAAATTTATTAAGACGCTTGAATGATGATGATTAAGAAAAAT 1452  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8781 GAATAACATTAAGTAAATACAGACAGAAATATAATAAAGAAAGAAATGATATTAATTA 8840  
QY 1453 GGATGATGATTTATGCGATTCCTACACCAATTA 1487  
DB 8841 AATATGATGTTTCAAGATTCAGAAAGCTATTAATTA 8875

RESULT 13  
US-08-676-967-2

Sequence 2, Application US/08676967  
Patent No. 5747317  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,967  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-676-967-2

Query Match 1.9%; Score 45.2; DB 1; Length 2277;  
Best Local Similarity 29.7%; Pred. No. 0.013;



Db 851 SNGARGARGAYWSNGAYYTNGARGARWSNGAYWSNATHGAYGAYGNGARGARYTNGCNC 910  
QY 519 CTCTGATATCATAGAGA 537  
::||:| :  
Db 911 ARWSNGAYACNWSNACNGA 929

Search completed: November 12, 2002, 15:30:55  
Job time : 141 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 14:47:29 ; Search time 74 Seconds  
(without alignments)  
11453.817 Million cell updates/sec

Title: US-08-961-083-55  
Perfect score: 2389  
Sequence: 1 TTCTTAGCGAGTGGACTGT.....TAAGTAAGCAAAATAAAC 2389

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_MA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PC01\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PC01\_NEM\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2388	100.0	2389	10	US-09-765-272-55
2	987.6	41.3	2290	10	US-09-765-272-65
3	385.4	16.1	1342	10	US-09-765-272-181
4	109	4.6	841	10	US-09-452-599-34
5	49.8	2.1	1959	10	US-09-864-761-4012
6	44.6	1.9	423	10	US-09-864-761-18355
7	44.6	1.9	487	10	US-09-864-761-1597
8	43.6	1.8	766	10	US-09-864-761-19608
9	43.6	1.8	1944	10	US-09-864-761-19262
10	42.4	1.8	660	10	US-09-864-761-19488
11	42.4	1.8	962	10	US-09-864-761-2772
12	42.4	1.8	4047	10	US-09-815-242-843
13	42.4	1.8	4050	10	US-09-815-242-9039
14	42	1.8	305	10	US-09-864-761-19262
15	42	1.8	496	10	US-09-864-761-2534
16	41.6	1.7	611	10	US-09-864-864-233
17	41.6	1.7	1845	10	US-09-864-864-313
18	41.6	1.7	1860	10	US-09-919-497-50
19	41.4	1.7	5361	9	US-09-742-096-2

20	41.4	1.7	6152	9	US-09-742-096-1	Sequence 1, Appl1
21	41.2	1.7	696	10	US-09-922-261-193	Sequence 193, App
22	41.2	1.7	699	10	US-09-922-261-191	Sequence 191, App
23	41.2	1.7	717	10	US-09-922-261-189	Sequence 189, App
24	41.2	1.7	774	10	US-09-922-261-187	Sequence 187, App
25	41.2	1.7	819	10	US-09-922-261-185	Sequence 184, App
26	41.2	1.7	1669	10	US-09-922-261-185	Sequence 184, App
27	40.8	1.7	2000	9	US-09-938-842A-4612	Sequence 4612, App
28	40.6	1.7	275	10	US-09-864-761-20595	Sequence 20595, A
29	40.2	1.7	475	10	US-09-864-761-1619	Sequence 1619, Ap
30	40	1.7	4316	10	US-09-880-107-3713	Sequence 3713, Ap
31	39.6	1.7	1458	10	US-09-815-242-4757	Sequence 4757, Ap
32	39.6	1.7	1599	10	US-09-815-242-4824	Sequence 8824, Ap
33	39.4	1.6	315	10	US-09-864-761-21723	Sequence 21723, A
34	39.2	1.6	462	10	US-09-864-761-1829	Sequence 3829, Ap
35	39.2	1.6	1074	10	US-09-861-451A-49	Sequence 49, Appl
36	39	1.6	477	10	US-09-864-761-5436	Sequence 5436, Ap
37	38.8	1.6	277	10	US-09-294-093B-1831	Sequence 1831, Ap
38	38.8	1.6	505	10	US-09-917-800A-314	Sequence 314, App
39	38.8	1.6	510	10	US-09-864-761-18737	Sequence 18737, A
40	38.8	1.6	2150	10	US-09-826-752-13	Sequence 13, Appl
41	38.2	1.6	1002	12	US-10-007-693-119	Sequence 119, App
42	38.2	1.6	2577	12	US-10-007-693-32	Sequence 32, Appl
43	37.8	1.6	184	10	US-09-864-761-20375	Sequence 20375, A
44	37.8	1.6	489	10	US-09-864-761-1607	Sequence 1607, Ap
45	37.8	1.6	583	10	US-09-864-761-20772	Sequence 20772, A

#### ALIGNMENTS

RESULT 1  
US-09-765-272-55  
Sequence 55, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
Applicant: Chai et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765, 272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961, 083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2389 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 55:  
US-09-765-272-55





|||||  
Db 2101 TAAGACTTCAAAAGCGATGAGAGAGCCAGTAGAGGAAACACCTCGAGCCACAACTGCC 2160  
Qy 2161 TCAGTAGAGACTGAAAAAGTAGAGCCCACTCAAAAGAGGAAAGTTTGGTTCGAA 2220  
|||||  
Db 2161 TCAGTAGAGACTGAAAAAGTAGAGCCCACTCAAAAGAGGAAAGTTTGGTTCGAA 2220  
Qy 2221 AGTAAGGATTCAGTCTGAAAGCCAAATGCAACAGAACTCTGCTGGTTTACGAATAA 2280  
|||||  
Db 2221 AGTAAGGATTCAGTCTGAAAGCCAAATGCAACAGAACTCTGCTGGTTTACGAATAA 2280  
Qy 2281 TTGACTCTTCAAAATTATGATTAACATATGATATCATGCGAAGAGAAATTTACTTGC 2340  
|||||  
Db 2281 TTGACTCTTCAAAATTATGATTAACATATGATATCATGCGAAGAGAAATTTACTTGC 2340  
Qy 2341 GTTGTAAAAAGAGTAATCTTCACTCTATCTGTAGTAAGAAAAAATTAAC 2389  
|||||  
Db 2341 GTTGTAAAAAGAGTAATCTTCACTCTGTAGTAAGAAAAAATTAAC 2389

## RESULT 2

US-09-765-272-65  
Sequence 65, Application US/09765272  
Patent No. US2002061545A1  
GENERAL INFORMATION:  
Applicant: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2290 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 65:  
US-09-765-272-65

Query Match 41.3% Score 987.6, DB 10; Length 2290;  
Best Local Similarity 67.7% Pred. No. 2,4e-240;  
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;  
Qy 1 TTCTTACGAGTGGAGCTATATCAAGCTGTAAGCGTTTGAAGAAA--TAAATCGTGTTC 57  
|||||  
Db 4 TTCTTACGAGTGGAGCTATATCAAGCTGTAAGCGTTTGAAGAAAAGTCTTAAATCGAGTTTC 63  
|||||  
Qy 58 CTATATAGATGAAAAACAAGCAGCAGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAG 117  
|||||

Db 64 TTATATAGATGATGATCAGGCTGTGTCAAAAAGCAGAAAACTTACACACATGATGAC 123  
Qy 118 CAGCGTGAAGGAATCAATGCTGAGCAAAATCGTATCAAGATTAACAGACCAGGCTATGT 177  
|||||  
Db 124 TAAGAGGAGGAGGATCAACGCCGCAACAAATNGTATCAAGATTAACGATCAAGATTATGT 183  
Qy 178 CACTTACATGGGAGCACCATCATATTATTCATATGTAAGGTTCTTTATGACCTATCAT 237  
|||||  
Db 184 GACCTCTATGAGAGACCATTAATCATTAATTAATGCAAGGTTCTTTATGATGCCATCAT 243  
Qy 238 CAGTGAAGATTTACTCATGAAGATCAAACTAATAGCTTAAAGATGAGAGATTTGTTAA 297  
|||||  
Db 244 CAGTGAAGAGCTCTCATGAAGATCAAACTAATAGCTTAAAGATGAGAGATTTGTTAA 303  
Qy 298 TGAAGTCAAGGCTGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGA 357  
|||||  
Db 304 TGAATCAAGGCTGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGA 363  
Qy 358 TGGTCCCAAGGAGATTAAGCTCCGTACAAAAGAGAAATCAATCGACAAAAGAGAGA 417  
|||||  
Db 364 TGCAGCTCATGCGGATTAATTTGCAACAAAAGAGATTAAACGTACAGAGAGAGAG 423  
Qy 418 TAGTCAACATCGTGAAGGTGGAACCTCCAGAAAGAGATGCTGTGCTTGGCAGCTTC 477  
|||||  
Db 424 CAGTCAATTAATCAT-----AATCAAGAGAGATTAATGCTGTGCTTGGCAGAGAGC 474  
Qy 478 GCAGAGACCTTATCTACAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGA 537  
|||||  
Db 475 CCAAGAGACCTTATACAAAGATGATGTTATATCTTCAATGATCATGATATCATAGAGA 534  
Qy 538 TACTGTGATGCTTATATGTTTCCATGATGATGATGATGATGATGATGATGATGATGAT 597  
|||||  
Db 535 CAGGCTGATGCTTATATGTTTCTTCAAGGAGCAGTCAATGATGATGATGATGATGATGAT 594  
Qy 598 GTTATCAGCTAGAGATGTTGCTGCTCAGAGAGCTTCTATCTGTGAGAAATGTGTC 657  
|||||  
Db 595 GTTATCAGCTAGAGATGTTGCTGCTCAGAGAGCTTCTATCTGTGAGAAATGTGTC 629  
Qy 658 AATTCAGAACCTTATCGCCGCAAAAATAGCATTAACCTTCAAGAACAACTGGGTACC 717  
|||||  
Db 630 -----ATTGAAATGAGAGACAGGAGATCGTCTCTTCAAGTTGATTAATGC 681  
Qy 718 TTCTGTAAACCAATCAAGCAATCAAAATCAACAAACCAACCAACAGCAACACTAGAC 777  
|||||  
Db 682 AATTCAGACTCAACCAAGATTTGTCAAGAACCAACCAATCTGACTGTATCTCAACTATCA 741  
Qy 778 TCAAGCAAGTCAAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
|||||  
Db 742 TCA---AATTCAGAGGGAAGAAACATTTCAAGCCCTTTAGCTGAATTTGTATGCTTAAACCTT 798  
Qy 838 GAGTCAAGCAATGTGAATCTGATGCTGCTTGTGTTGATCCAGCAACAATCAACAAGTCG 897  
|||||  
Db 799 ATCAGAACCCCATGTGGAATCTGATGCTGCTTATTTTCGACCAAGCCCAATATCAAGTCG 858  
Qy 898 AACAGCTAGAGGTGTGAGTGGCCACAGGAGATCAATTCACACTTATCCCTTACTCTCA 957  
|||||  
Db 859 AACCCGACAGAGGTGATGCTGCTTATGATGATGATGATGATGATGATGATGATGATGAT 918  
Qy 958 AATGCTGAATGGAAGAAAGCAATGCTGATTAATTTCCCTTCTGTTATCTGTTCAACCA 1017  
|||||  
Db 919 AATGCTGAATGGAAGAAAGCAATGCTGATTAATTTCCCTTCTGTTATCTGTTCAACCA 978  
Qy 1018 TTGGGTACCAAGTTTCAAGGCGCAAGAACCAAGTCCACACACGACTCGGAACTATGTC 1077  
|||||  
Db 979 TTGGGTACCAAGTTTCAAGGCGCAAGAACCAAGTCCACACACGACTCGGAACTATGTC 1038  
Qy 1078 AGGCCGCAAGCTGACCAAAATCTTAAATATAGCAATTCCTC-----TTT 1125  
|||||  
Db 1039 AAGTCGCAAGCTGACCAAAATCTTCAACCAAGTCCACACCAATTCGAATGATGAGAAAT 1098  
Qy 1126 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1185  
|||||  
Db 1099 GGTCAAGAGAGCTGTTCAAGAAAGTAGGCGATGTTATGATGTTGAGAGAAATGAGATTTTC 1158



Db 436 TAATGTCCTGAGCAAGAGTCTCAGGAGCGAATATACGACCAATATGATGGTTATGTCCTTTAA 495  
 Oy 517 TGCTTCTATATCATAGAGGACTACTGGTGGATGCTTATATTCGTTCTCATGAGATCATTA 576  
 Db 496 TCCACCTATATATTCGAGAGATAGCGGTAAGCTTATATTCGTTCTCATGAGAGGTCACCTA 555  
 Oy 577 CCATTACATATCTAAGATATGATGATTCAGTACAGGAGATGGCGTGGTCGAGAAAGCCCTTCCT 636  
 Db 556 TCACCTACATATCCCAAAAGCGATTTATCTGCTAGTGAATATAGCAGCAGCTAAAGCACATCT 615  
 Oy 637 ATCTGTCGAGGAAATCTGTCAAAATTCAGAAACCTATGCGCCACAAATATAGCGATTAACAC 696  
 Db 616 GCGTCGAAAAAATATATGCAACGAGCAGTATTAAGCTATATCTTCAACAGCAGCTGAGACAA --- 672  
 Oy 697 TTCAGAACCAAACTGGGTACTCTTCTGTAAAGCAATCCAGGAACACTACAATTAACACAAAG 756  
 Db 673 -----TAAACGCGCATATCTGT 687  
 Oy 757 CAACACAGCAACACTAATCAGTCAAGCAAGTCAAGTAAATGACATTTGATGCTCTTGAA 816  
 Db 688 AGCAAAAGATCACTACTACAGCAAGCCAGCAATTAATCTCAAAATCTCCAGAGTCTTTTAA 747  
 Oy 817 ACAGCTCTACAAACTGCGCTTTGATGTCAAAGCAGCATGTAGATATCTGATGGCCCTGTCTTGA 876  
 Db 748 GGAATCTATATATTCACCTAGCGCCCAACGTTACAGTAAATACAGATGGCTGTGTTTGA 807  
 Oy 877 TCCACCACAAAATCCACAGTTCGAAAGCTGATAGAGTGTGACAGTCCACAGGAGATCATTA 936  
 Db 808 CCTGCTTAGATTTATCACTACTGTCACCCAAATGGAATGCGATTCGCAATGGGACCATTA 867  
 Oy 937 CCACCTTCATCCCTTACTCTCAAAATGCTGAAATGGGAAGACGAATGCGCTGATTAATTC 996  
 Db -868 CCACCTATATCTCTTACAGCAAGCTTTCTGCTTAGAAGAAAAAGATTGCCAGATGTCGCC 927  
 Oy 997 CCTTCCT 1003  
 Db 928 TATCAGT 934  
  
 RESULT 4  
 US-09-452-599-34/c  
 : Sequence 34, Application US/09452599  
 : Patent No. US20020055101A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Bergeron, Michel G.  
 : APPLICANT: Ouellette, Marc  
 : APPLICANT: Roy, Paul H.  
 : TITLE OF INVENTION: Specific and Universal Probes and Amplification Primers  
 : TITLE OF INVENTION: to Rapidly Detect and Identify Common Bacterial  
 : TITLE OF INVENTION: Pathogens and Antibiotic Resistance genes from Clinical  
 : TITLE OF INVENTION: Specimens for Routine Diagnosis in Micro  
 : FILE REFERENCE: 12287.31  
 : CURRENT APPLICATION NUMBER: US/09/452.599  
 : CURRENT FILING DATE: 1999-12-01  
 : PRIOR APPLICATION NUMBER: 08/526.840  
 : PRIOR FILING DATE: 1995-09-11  
 : PRIOR APPLICATION NUMBER: 08/304.732  
 : PRIOR FILING DATE: 1994-09-12  
 : NUMBER OF SEQ ID NOS: 177  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 34  
 : LENGTH: 841  
 : TYPE: DNA  
 : ORGANISM: Streptococcus pneumoniae  
 US-09-452-599-34

	Query Match	468;	Score 109;	DB 10;	Length 841;
	Best Local Similarity	65.4%;	Pred.No. 4.3e-18;		
	Matches 176; Conservative	0;	Pmismatches 90;	Indels	3;
				Gaps	1.
Oy	2090 GAAGATCCCAATTAAACTTCGAACCGCATGTAGAACCACCTGGTGAG	2149			
I I		I I I I	I I		

Db	633	GAGTCGCCAAAACCAACAGAGGAACCGAAGAAAGATACACAGAAAGATACACAGAGGA	574
Qy	2150	CCAGAATGCTCTCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAGACAGAACTT	2209
Db	573	TCAGAAAGAACTCGATGTCGAGACTGAAAGGTATGAAGAA--AACTGAGAGAGCTGAAGAT	517
Qy	2210	TTGCTGTGGAGAGTAGACGGAATTCAGTCTAAAGCCCAATGCAACAGAACTCTAGCTGT	2268
Db	516	TTACTTGGAAAAATCCAGATCCCATTTATCAATCCATATGCCAAGAGAGCTCTCACAGA	457
Qy	2270	TTACGAATAATTTGACTCTTCAAATTATGATTAACATATGATCTATGCGCAGAGACAGA	2329
Db	456	TTAAAAAATAATTACTATTGTCACCCAGCAACACAACTACTATTATGCGCAGAACTGAA	397
Qy	2330	AAATTACTGCTGCTGTAAAGAGATAA	2358
Db	396	AAACTATTGGCTTATTATTAAGAGAGTAA	368

```

RESULT 5
US-09-864-761-4012
; Sequence 4012, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4012
;
; LENGTH: 1959

```

```

: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL022334.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
: US-09-864-761-4012

Query Match      2.1%; Score 49.8; DB 10; Length 1959;
Best Local Similarity 50.5%; Pred. No. 0.0063;
Matches 148; Conservative 0; Mismatches 142; Indels 3; Gaps 1;

QY 279 AAGATGAGATATTGTAATGAGGTCAAGGTGATGTTATCAAGTAGTGAATAAT 338
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 537

QY 339 ACTATGTTACTTAAGATGCTGCCACGCGATTAACGTCCTGACAAAGAGAAATCA 398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 538 ATGGTATGATGGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 597

QY 399 ATGACAAAACAAAGACATGACATGCTGAAGCTGCACTCCAGAAAGCATGTG 458
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 598 ATATGAGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 657

QY 459 CTGTGCTTGGACAGCTTCCCAAGACGCTATCTACAGATGATGATGATGATGATG 518
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 658 GTGATGATGGGAAGATG---ATGACAGTATGATGATGATGATGATGATGATGATG 714

QY 519 CTCTGATATCATAGAGATAGTGTGATGCTTATATGTTCTCATGAGAT 571
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 715 GGGATGATGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 767

RESULT 6
US-09-864-761-18355
: Sequence 18355, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aomicia-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
```

```

: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 18355
: LENGTH: 423
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC010133.1
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
: US-09-864-761-18355

Query Match      1.9%; Score 44.6; DB 10; Length 423;
Best Local Similarity 51.8%; Pred. No. 0.062;
Matches 101; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 2012 CACCTGACGACGTCACATTCATATGATGATGGGCCAATGCCAGTAGCATGTGTTA 2071
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7 CTCTCAGAACATGCGAGTATCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGATGGAG 66

QY 2072 GGCAGAAAGACCAAGTGAAGATCCAAATAGAACTTCAAGCGGATGAAGGCCAGTA 2131
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 126

QY 2132 GAGGAAGACCTGAGCCAGAGATCCCTCAAGTAGAGACTGAGAGAGAGAGAGAGAG 2191
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 GACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 186

QY 2192 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 2206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 CACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 201

RESULT 7
US-09-864-761-1597
: Sequence 1597, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: Aomicia-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
```

```

PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 1597
LENGTH: 487
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010133.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
US-09-864-761-1597
```

```

Query Match          1.9%; Score 44.6; DB 10; Length 487;
Best Local Similarity 51.8%; Pred. No. 0.066; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 94;
```

```

QY 2012 CACCTGAGCAAGTCACATCTATGATGATGGGCAATGCCAGTACGATGTGA 2071
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 166 CTCTGAGCAATGCGATGATCTGTAAGAGAGGAGAGAGAGAGAGAGAGAG 225
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2072 GCGAAGAAGACCAAGTGAAGTCCAAATTAAGAACTTCAAGCGGATGAAGCCAGTA 2131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 226 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2132 GAGGAACACCTGCTGAGCCAGAACTCCCTCAAGTAGACAGCTGAAAAGTAAAGCCCAA 2191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 286 GAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2192 CTCAGAAGACAGAA 2206
```

DB 346 CACAAGAGAGAGAA 360

RESULT 8  
US-09-864-761-19608/c  
Sequence 19608, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aecmics-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 19608

LENGTH: 766

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AL008720.1

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2

OTHER INFORMATION: SWISSPROT HIT: P17164, EVALU 4.10e+00

OTHER INFORMATION: EST\_HUMAN HIT: AW844901.1, EVALU 2.90e+00

US-09-864-761-19608

Query Match 1.8%; Score 43.6; DB 10; Length 766;

Best Local Similarity 46.4%; Pred. No. 0.15;  
Matches 180; Conservative 0; Mismatches 204; Indels 4; Gaps 1;

```
OY 186 ATGGCGACCCCTTCATATATACATAGTTCCTTTCAGCGCTATCATCAGGAG 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 ATGAAAATGATGATGACGGGATGATGAGATGATATATGATGATGATGATG 370
OY 246 AATTACTCATGAAAGATCCAACTATAGCTAAAGATGAGATATGTTATGAGCTCA 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 ATTATGATATGATGATGAGGAGATGAGAGATGAGATGATGATGATGATGAT 310
OY 306 AGGGTGATATGTTATCAAGCTAGATGAAATATCATATGTTTACCTTAAGAGTCTGCC 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 ---TGATGGTATGAGAGGAGATGGCGATGATGATGATGATGATGATGATG 254
OY 366 ACCGGATTAACGTCCTACAAAGAGAAATCATGACGACAAAACAAGCATATGCAAC 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 ATGAAGATGACGATGATGAGAGAGAGGTTGGTGTGATGATGATGATGATGATG 194
OY 426 ATCGTAAGTGGGAACTCCAAAGAAAGATGCTGTTCCTTGGCAGCTTCCCAAGGAC 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 GTGTGATGATGATGATGAAAGATGATGATGATGATGATGATGATGATGATG 134
OY 486 GCTATCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 TTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 74
OY 546 ATGCTTATATCGTTCCTCATGAGATCA 573
Db 73 AAGAGATGATGATGATGATGATGATG 46
```

## RESULT 9

```
US-09-864-761-2825/C
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 2825
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825
```

Query Match 1.8%; Score 43.6; DB 10; Length 1944;

Best Local Similarity 46.4%; Pred. No. 0.23;  
Matches 180; Conservative 0; Mismatches 204; Indels 4; Gaps 1;

```
OY 186 ATGGCGACCATATCATATTAATCAATGTAAGTTCCTTATGACGCTATCATCAGTGAAG 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 ATGAAAATGATGATGACGGTATGATGATGATGATGATGATGATGATGATGATG 607
OY 246 AATTACTCATGAAAGATCCAACTATAACCTAAAGATAGATATGTTTATGAGCTCA 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 ATTATGATATGATGATGATGAGGAGATGAGAGATGATGATGATGATGATGATG 547
OY 306 AGGGTGATATGTTATCAAGCTAGATGAAATATCATATGTTTACCTTAAGAGTCTGCC 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 ---TGATGGTATGAGAGGAGAGATGGCGATGATGATGATGATGATGATGATG 431
OY 366 ACCGGATTAACGTCCTACAAAGAGAAATCATGACGACAAAACAAGCATATGCAAC 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 ATGAAGATGACGATGATGAGAGAGAGGTTGGTATGATGATGATGATGATGATG 431
OY 426 ATCGTAAGTGGGAACTCCAAAGAAAGATGCTGTTCCTTGGCAGCTTGGCAAGGAC 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 371
OY 486 GCTATCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 TTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 311
OY 546 ATGCTTATATCGTTCCTCATGAGATCA 573
Db 310 AAGAGATGATGATGATGATGATGATG 283
```

## RESULT 10

```
US-09-864-761-19488
; Sequence 19488, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	1.8%	Score 42.4	DB 10	Length 660
	Best Local Similarity	48.7%	Pred. No. 0.28		
	Matches 146	Conservative 0	Mismatches 151	Indels 3	Gaps 1
QY	278	AAAGATGAGGATATGTTAATGACGCTCAAGGCGTGATATGTTATCAAGTAGATGAAAA	337		
Db	108	AATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	167		
QY	338	TACATATGTTTACCTTAAGGATGCTGCCACGCGGATTAACGTCGGTACAAAAGGAAATC	397		
Db	168	GGTATATGGTGGTGAATGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT	227		
QY	398	AATGACGAAAAACAAGACATAGTCGAACATCGTGAAG--TGGAACCTCCAGAAACAAGT	454		
Db	228	GATCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	287		
QY	455	GGTGCTGTTGCCCTTGGACAGTTTGGCAAGACGCTTACTACAGATGATGATATATCTTT	514		

```

Db      288  AGTATGATGATGTTATGAGCAGTACCATGCTGATGTATGATGATGATGTTATGGCAGT 347
Oy      515  AATGCTTCTGATATCATAGAGGACTGAGTATGCTTATATCGTTCCATCGAGAGATCAT 574
Db      348  GACGATGCTGATGCTGATGACGATGCTGTGATGATGCTGATGATGCTGATGCTGATGATCAT 407

RESULT 11
US-09-864-761-2772
: Sequence 2772, Application US/09864761
: Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2772.
LENGTH: 962
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008125.9
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5

```







;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 2534  
;; LENGTH: 496  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC011416.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
US-09-864-761-2534

Query Match 1.8%; Score 42; DB 10; Length 496;  
Best Local Similarity 52.9%; Pred. No. 0.3;  
Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 2038 TGATGATGGGCAATGCCAGTGTGTAGGCAAGAACCCACAGTGAAGATCC 2097  
DB 484 TGGTTGGGGGTGGGAGAGAGAGCGAGAAAGAGAGAGAGAGAGAGAGAGA 425  
QY 2098 AATAAGAACTTCAAAGCGATGAGAGCCAGTGAAGAAACACTGAGCCAGAGT 2157  
DB 424 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 365  
QY 2158 CCTCAAGTAGAGCTGAAGAAAGTAGAAGCCCAACTCAAGAAAGCAGAG 2207  
DB 364 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAG 315

Search completed: November 12, 2002, 15:28:40  
Job time : 101 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	4163	100.0	796	19	AAW55090	Streptococcus pneumoniae
2	4163	100.0	796	19	ABP54584	S. pneumoniae SP03
3	4163	100.0	819	21	AAAB01468	Recombinant variant
4	3218	77.3	821	21	AAAB12727	Streptococcus pneumoniae
5	3218	77.3	821	21	AAAB12766	Streptococcus pneumoniae
6	3218	77.3	821	23	AAU084026	truncated variant
7	3218	77.3	840	21	AAAB12710	Streptococcus pneumoniae
8	3218	77.3	840	23	AAU75933	Streptococcus pneumoniae
9	2961	71.1	690	23	AAAB12745	Streptococcus pneumoniae
10	2961	71.1	690	23	AAU084093	truncated variant

45	1666	40.0	334	21	AA812728
44	1795.5	43.1	1378	23	AA814053
43	1795.5	43.1	1139	23	AA804055
42	1795.5	43.1	568	23	AA804088
41	1795.5	43.1	568	21	AA812741
40	1808	43.4	612	23	AA804031
39	1888	45.3	555	23	AA804047
38	1888	45.3	555	21	AA812734
37	1891.5	45.4	1265	23	AA804057
36	1898	45.6	1126	23	AA804058
35	2170	52.1	613	23	AA804279
34	2170	52.1	613	21	AA812730
33	2431.5	58.4	721	20	AA805753
32	2649.5	63.6	763	23	AA805489
31	2649.5	63.6	763	19	AA805095
30	2713	65.1	816	21	AA812753
29	2714	65.2	816	21	AA812758
28	2720	65.3	811	21	AA812760
27	2720	65.3	811	21	AA812757
26	2727	65.5	816	21	AA812756
25	2728	65.5	816	21	AA812759
24	2733	65.6	834	23	AA812758
23	2750.5	66.0	838	23	AA812750
22	2750.5	66.0	819	23	AA812720
21	2750.5	66.0	819	23	AA812754
20	2750.5	66.0	819	21	AA812763
19	2750.5	66.0	819	21	AA812760
18	2760	66.3	811	21	AA812763
17	2771	66.5	807	21	AA812765
16	2772	66.6	838	21	AA801466
15	2772	66.6	820	21	AA812756
14	2786	66.9	805	21	AA812763
13	2786	66.9	819	21	AA801469
12	2795	67.0	819	21	AA812766
11	2812	67.5	826	21	AA811939
10	2812	67.5	826	21	AA811939

[illegible]

	RESULT
AAW55090	1
ID	AAW55090 standard; Protein; 796 AA.
XX	
AC	AAW55090;
XX	
DT	02-OCT-1998 (first entry)
XX	
DE	Streptococcus pneumoniae SP0036 protein.
XX	
KW	Streptococcus pneumoniae; antigen; vaccine; in detection; pneumonia; otitis media; meningitis
XX	
OS	Streptococcus pneumoniae.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 456
FT	/label= unknown
FT	/note= "encoded by GNA"
XX	
PN	W09818930-A2.
XX	
PD	07-MAY-1998.
XX	
PE	30-OCT-1997; 97WO-US19422.
XX	
PR	31-OCT-1996; 96US-0029960.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Choi GH, Bromocky J A, Johnson LS, Kunsch CA
DR	WPI; 1998-272224/24.

DR N-PSDB; AAN27351.  
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
XX  
PS Claim 11; Page 59-60; 118pp; English.  
XX  
CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
XX  
XX

SQ Sequence 796 AA:

Query Match 100.0%; Score 4163; DB 19; Length 796;  
Best Local Similarity 100.0%; Pred. No. 2.7e-299;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEOIVIKITDGGYVT 60  
DB 1 STELGILQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEOIVIKITDGGYVT 60  
OY 61 SHGDHYHYNGKVPYDAIISEELMKDPYKIKDEDIVNEVGKGYIKYDGKYYVYLKDA 120  
DB 61 SHGDHYHYNGKVPYDAIISEELMKDPYKIKDEDIVNEVGKGYIKYDGKYYVYLKDA 120  
OY - 61 SHGDHYHYNGKVPYDAIISEELMKDPYKIKDEDIVNEVGKGYIKYDGKYYVYLKDA 120  
DB - 61 SHGDHYHYNGKVPYDAIISEELMKDPYKIKDEDIVNEVGKGYIKYDGKYYVYLKDA 120  
OY 121 AHADNVRTEEINROKQEHSHQREGTPRNDGAVALARISQGRYTTDGYIFNADIIEDT 180  
DB 121 AHADNVRTEEINROKQEHSHQREGTPRNDGAVALARISQGRYTTDGYIFNADIIEDT 180  
OY 181 GDAYIYPHGDHYHYIPKNLSASFLAFAFLSGRNLNSFRYRQNDNSTRMWWVS 240  
DB 181 GDAYIYPHGDHYHYIPKNLSASFLAFAFLSGRNLNSFRYRQNDNSTRMWWVS 240  
OY 241 VSNPCTNTNTNSNTNSQASQNDISLKLQYKLPLOSRRHVESDGLVFPQAQITSTR 300  
DB 241 VSNPCTNTNTNSNTNSQASQNDISLKLQYKLPLOSRRHVESDGLVFPQAQITSTR 300  
OY 301 ARGVAAPPHGDHYHFIPYSQMSLEERIALIPLRYRSNHWPDSRPEQSPQTPPEPSFG 360  
DB 301 ARGVAAPPHGDHYHFIPYSQMSLEERIALIPLRYRSNHWPDSRPEQSPQTPPEPSFG 360  
OY 361 POPAPPLKIDSNSISQVLRKVGEGYVEEKGISRYVAKOLPSTYVNLKSKQSS 420  
DB 361 POPAPPLKIDSNSISQVLRKVGEGYVEEKGISRYVAKOLPSTYVNLKSKQSS 420  
OY 421 VSHITLAKKENVAPRDOEFYDKAYNLTLEAHKALFXNKRNSDFQALDLERLNDSTN 480  
DB 421 VSHITLAKKENVAPRDOEFYDKAYNLTLEAHKALFXNKRNSDFQALDLERLNDSTN 480  
OY 481 KEKLVDDLALAPLTIHPERLGRKPSQIEYTEDEVRIALQADKYTSDGIFDEHDIISD 540  
DB 481 KEKLVDDLALAPLTIHPERLGRKPSQIEYTEDEVRIALQADKYTSDGIFDEHDIISD 540  
OY 541 EBDAYVTTPMGSHWIGKSLSDKEVAAQATYKKEGILPPSPDADVKANPPGDSAAIY 600  
DB 541 EBDAYVTTPMGSHWIGKSLSDKEVAAQATYKKEGILPPSPDADVKANPPGDSAAIY 600  
OY 601 NNVKGEKKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIRKAWPDDHRYKAPNGYTLED 660  
DB 601 NNVKGEKKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIRKAWPDDHRYKAPNGYTLED 660

OY 661 LEATIKYVEHDEPRHSDNGWGNASEHYLGKDKHSEDPNKNFKADEPVEETPAPEVP 720  
DB 661 LEATIKYVEHDEPRHSDNGWGNASEHYLGKDKHSEDPNKNFKADEPVEETPAPEVP 720  
OY 721 QVETKEVNEQLKEAVYLLAKYTDSSIKANATETLAGLRNLTLOIMDNNSITMAEAEKLLA 780  
DB 721 QVETKEVNEQLKEAVYLLAKYTDSSIKANATETLAGLRNLTLOIMDNNSITMAEAEKLLA 780  
OY 781 LKGSNPSVSKKKTIN 796  
DB 781 LKGSNPSVSKKKTIN 796

RESULT 2  
ABP54584  
ID ABP54584 standard; Protein; 796 AA.

XX ABP54584;

XX 04-SEP-2002 (first entry)

XX S. pneumoniae SP036 protein sequence SEQ ID NO:56.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

XX antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-0765272.

XX 30-OCT-1997; 97US-0961083.

XX (CHOI/) CHOI G. H.

XX (KUNS/) KUNSCH C. A.

XX (BARA/) BARASH S. C.

XX (DILL/) DILLON P. J.

XX (DOUG/) DOUGHERTY B.

XX (FANN/) FANNON M. R.

XX (ROSE/) ROSEN C. A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;

XX WPI: 2002-479261/51.

XX N-PSDB; ABQ84819.

XX New Streptococcus pneumoniae antigens, useful for detecting  
PT Streptococcus and for preventing or attenuating disease caused by  
PT Streptococcus infection -  
XX

XX Claim 11; Page 27; 70pp; English.

XX ABQ84792 to ABQ84904 represents nucleic acids which encode the  
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54659.XX The S. pneumoniae antigens have antibacterial activity and can be  
CC used in vaccines. The S. pneumoniae antigens can also be used to  
CC prevent or attenuate a Streptococcal infection in an animal. The  
CC polynucleotides encoding the S. pneumoniae antigens can be used to  
CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
XX which are used in an example from the present invention.

XX Sequence 796 AA;

XX

Query Match 100.0%; Score 4163; DB 23; Length 796;  
Best Local Similarity 100.0%; Pred. No. 2.7e-299;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEOIVIKITDGGYVT 60

|||||  
Db 1 SYELGLYQARTYKNNRNSYIDGKATQKTENLTDEVSKREGINAEOIVIKITDQGYT 60  
OY 61 SHGDHYHYNGKVPYDAITSEELMKDPYKDKDEDIVNEVGVIKVDGYYVYLKDA 120  
Db 61 SHGDHYHYNGKVPYDAITSEELMKDPYKDKDEDIVNEVGVIKVDGYYVYLKDA 120  
OY 121 AHADVNRTEEINRQKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180  
Db 121 AHADVNRTEEINRQKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180  
OY 181 GDAYIVPHGDHYHYIPKNELASASELAAPAFISGRGNLSNSRTYRRQNSDNTSRFMWVS 240  
Db 181 GDAYIVPHGDHYHYIPKNELASASELAAPAFISGRGNLSNSRTYRRQNSDNTSRFMWVS 240  
OY 241 VSNPGTNTNTNSNTNSQASQSDNDISLLKQYKPLSQOHVESDGLVFPDAQITST 300  
Db 241 VSNPGTNTNTNSNTNSQASQSDNDISLLKQYKPLSQOHVESDGLVFPDAQITST 300  
OY 301 ARGVAVPHGDHYHYIPYSQMSSELEERARIIPLRYSNNHWVDSRPEQSPQTPPEPSFG 360  
Db 301 ARGVAVPHGDHYHYIPYSQMSSELEERARIIPLRYSNNHWVDSRPEQSPQTPPEPSFG 360  
OY 361 POPAPNLKIDSNSLSVQLVRRVGEVYVEEKGISRYVFAKDLPSFTYKNLESKLSKQES 420  
Db 361 POPAPNLKIDSNSLSVQLVRRVGEVYVEEKGISRYVFAKDLPSFTYKNLESKLSKQES 420  
OY 421 VSHHTLAKKENVAPRDOEFYDKAYNLTPFAHKALFYKNGRNSDFQALDKLERLNDSTN 480  
Db 421 VSHHTLAKKENVAPRDOEFYDKAYNLTPFAHKALFYKNGRNSDFQALDKLERLNDSTN 480  
OY 481 KEKLVDDLAFAPITTHPERLGKPNQSOIEYTEDEYVRIQLADKYTSDGYIFDEHDIISD 540  
Db 481 KEKLVDDLAFAPITTHPERLGKPNQSOIEYTEDEYVRIQLADKYTSDGYIFDEHDIISD 540  
OY 541 EBDAYVTTPMGSHWIGKDSLSDKERVAQAAYTKEGILIPSPDADVKANPTGDSAAIY 600  
Db 541 EBDAYVTTPMGSHWIGKDSLSDKERVAQAAYTKEGILIPSPDADVKANPTGDSAAIY 600  
OY 601 NNVKGKRRIPVRLPYMVEHYEVKNGNLIPHKDHYHNKFAWDDHRYKAPNGYTTLED 660  
Db 601 NNVKGKRRIPVRLPYMVEHYEVKNGNLIPHKDHYHNKFAWDDHRYKAPNGYTTLED 660  
OY 661 LEATIKYVEHDERPHSNDGNGNASEHYLGKKHSEDPNKNFKADEEVEETPAPEVP 720  
Db 661 LEATIKYVEHDERPHSNDGNGNASEHYLGKKHSEDPNKNFKADEEVEETPAPEVP 720  
OY 721 QVETEKVEAQLKEAEVLLAKYTDSSILKANATEFLAGLRNNLTLOIMDNNISMAAEKLLA 780  
Db 721 QVETEKVEAQLKEAEVLLAKYTDSSILKANATEFLAGLRNNLTLOIMDNNISMAAEKLLA 780  
OY 781 LKGSNPSSVSKEKIN 796  
Db 781 LKGSNPSSVSKEKIN 796

RESULT 3

ID AAB01468 standard: Protein: 819 AA.

XX AAB01468;

XX 20-OCT-2000 (first entry)

XX Recombinant variant of Sp36 (Sp36a) of S. pneumoniae.

XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;

KM histidine triad residue; Sp36; antibody; otitis media;

KM nasopharyngeal infection; bronchial infection; bronchitis; sepsis;

XX meningitis; lobar pneumonia.

XX Streptococcus pneumoniae.

FH Key Location/Qualifiers  
FT Region 63..68  
FT Region /label= Histidine triad residue 118..145  
FT Region /label= Coiled coil region 189..194  
FT Region /label= Histidine triad residue 309..314  
FT Region /label= Histidine triad residue 406..434  
FT Region /label= Coiled coil region 462..493  
FT Region /label= Coiled coil region 550..555  
FT Region /label= Histidine triad residue 634..639  
FT Region /label= Histidine triad residue 724..751  
FT Region /label= Coiled coil region

PN W0200037105-A2.

PD 29-JUN-2000.

XX 21-DEC-1999; 99WO-US30390.

XX 21-DEC-1998; 98US-0113048.

PR (MEDI-) MEDIMUNE INC.

PI Johnson LS, Koenig S, Adamou JE;

DR WPI: 2000-452129/39.

DR N-PSDB: AAA47604.

PT Vaccine useful for prophylaxis and treatment of pneumococcal infections

PT such as otitis media, nasopharyngeal and bronchial infections,

PT comprises Streptococcus pneumoniae proteins

PS Claim 1: Page 61-64; 70pp; English.

XX Although a number of proteins have been suggested as being involved  
CC in the pathogenicity of Streptococcus pneumoniae, there still remains  
CC a need to identify polypeptides having epitopes in common from  
CC various strains of S. pneumoniae in order to utilise such  
CC polypeptides in vaccines to protect against a wide variety of  
CC S. pneumoniae. New vaccine compositions are described which comprise a  
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino  
CC acids in length that comprise at least one histidine triad residue  
CC (HxxHxx) or a coiled-coil region, or an antibody directed against  
CC these features. The vaccine is useful in protecting against infection  
CC by Streptococcus pneumoniae. The vaccine composition comprising  
CC antibodies to is useful for passive immunization for treating  
CC pneumococcal infections which includes otitis media, nasopharyngeal  
CC and bronchial infections.

XX Sequence 819 AA;

Query Match 100.0%; Score 4163; DB 21; Length 819;

Best Local Similarity 99.9%; Pred. No. 2.9e-299;

Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYELGLYQARTYKNNRNSYIDGKATQKTENLTDEVSKREGINAEOIVIKITDQGYT 60

Db 21 SYELGLYQARTYKNNRNSYIDGKATQKTENLTDEVSKREGINAEOIVIKITDQGYT 80

OY 61 SHGDHYHYNGKVPYDAITSEELMKDPYKDKDEDIVNEVGVIKVDGYYVYLKDA 120

Db 81 SHGDHYHYNGKVPYDAITSEELMKDPYKDKDEDIVNEVGVIKVDGYYVYLKDA 140

OY 121 AHADVNRTEEINRQKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180

Db 141 AHADVNRTEEINRQKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 200

```
QY 181 GDAYIVPHGDHYHYIPKNELSASELAFAEFLSGRGLNSRRTYRRONSNDTSRTNWVPS 240
   |||
Db 201 GDAYIVPHGDHYHYIRKNELSASELAFAEFLSGRGLNSKRTRYRRONSNDTSRTNWVPS 260
QY 241 VSNPGTTNTNTSNNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGLVEFPAQITSRT 300
   |||
Db 261 VSNPGTTNTNTSNNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGLVEFPAQITSRT 320
QY 301 ARGVAAPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQTPPEPSPG 360
   |||
Db 321 ARGVAAPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQTPPEPSPG 380
QY 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEBEKGISRYVEAKDLPSFTVKMLESKLSKQES 420
   |||
Db 381 POPAPNLKIDSNSLSVSQLVRKVGEGYVEBEKGISRYVEAKDLPSFTVKMLESKLSKQES 440
QY 421 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFYXNGKNSDFQALDKLERLNDESTN 480
   |||
Db 441 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFEKNGKNSDFQALDKLERLNDESTN 500
QY 481 KEFLVDDLAFLAPITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIFDEHDIISD 540
   |||
Db 501 KEFLVDDLAFLAPITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIFDEHDIISD 560
QY 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKKEGILPPSPDADYKANPTGDSAAATY 600
   |||
Db 561 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKKEGILPPSPDADYKANPTGDSAAATY 620
QY 601 NRVKGEKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGYTLDE 660
   |||
Db 621 NRVKGEKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGYTLDE 680
QY 661 LEFTITYYVHPDERHSNDGMSNASEHYLGKDHSEDPKKNKADBEPEETPAPEVP 720
   |||
Db 681 LEFTITYYVHPDERHSNDGMSNASEHYLGKDHSEDPKKNKADBEPEETPAPEVP 740
QY 721 OVETEKEVAOLKEAEVLLAKVTDSILKANATEYLAGLRNNLTQIMDNNSIMAFAEKLLA 780
   |||
Db 741 OVETEKEVAOLKEAEVLLAKVTDSILKANATEYLAGLRNNLTQIMDNNSIMAFAEKLLA 800
QY 781 LKGSNPPSSYSKEKIN 796
   |||
Db 801 LKGSNPPSSYSKEKIN 816

RESULT 4
AAB12727
ID AAB12727 standard; Protein; 821 AA.
XX
AC AAB12727;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-11M protein antigen seq. ID NO:60.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
```

```
XX
DR MPI: 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Claim 18; Fig 25; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-11M protein antigen.
XX
SQ Sequence 821 AA:
XX
Query Match 77.3%; Score 3218; DB 21; Length 821;
Best Local Similarity 75.2%; Pred. No. 2,8e-229;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
QY 1 SYELGLYQARTYKENNRVSYIDGKQATOKTENLTPEVSKRGINAEOIVIKITPOGYVT 60
   :||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2 AYELGLHQQYTKENNRVSYIDGKQATOKTENLTPEVSKRGINAEOIVITPDGYVT 61
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDDEIVNEVGKGYIVKGYVYVYLDA 120
   |||
Db 62 SHGDHYHYNGKVPYDAIISEELMKDPNYQKDSQDIVEIRKGYIVKGYVYVYLDA 121
QY 121 AHADNVRKTEELNRQKQESHQREBGTFRNDGAVALARSQGRYTTDDGYIFNADSIIDET 180
   |||
Db 122 AHADNVRKTEELNRQKQESHQREBGTFRNDGAVAFARQGRYTTDDGYIFNADSIIDET 181
QY 181 GDAYIVPHGDHYHYIPKNELSASELAFAEFLSGRGLNSRRTYRRONSNDTSRTNWVPS 240
   |||
Db 182 GDAYIVPHGDHYHYIPKNELSASELAFAEFLSGRGLNSLRTYRRONSNDTSRTNWVPS 241.
QY 241 VSNPGTTNTNTSNNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGLVEFPAQITSRT 300
   |||
Db 242 VSNPGTTNTNTSNNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGLVEFPAQITSRT 301
QY 301 ARGVAAPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQTPPEPSPG 360
   |||
Db 302 ARGVAAPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQTPPEPSPG 361
QY 361 POPAPNLK-IDSN---SLVSQLVRKVGEGYVEBEKGISRYVEAFKDLPSFTVKMLESKLS 416
   |||
Db 362 POPAPNLK-IDSN---SLVSQLVRKVGEGYVEBEKGISRYVEAFKDLPSFTVKMLESKLS 421
QY 417 KOESVSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFYXNGKNSDFQALDKLERLND 476
   |||
Db 422 KOESVSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFYXNGKNSDFQALDKLERLND 481
QY 477 ESTNNKELVDDLAFLAPITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIFDEHD 536
   |||
Db 482 VSDKVLVDDLAFLAPITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIFDEHD 541
QY 537 IISDEGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKKEGILPPSPDADYKANPTGDSAA 596
   |||
Db 542 IISDEGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKKEGILPPSPDADYKANPTGDSAA 601
QY 597 AAIYNRKGEKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGY 656
   |||
Db 602 AAIYNRKGEKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGY 661
QY 657 TLEDEAFATIKYVVEHPDERHSNDGMSNASEHYLGKDHSEDPKKNKAD----- 706
   |||
Db 662 TLEDEAFATIKYVVEHPDERHSNDGMSNASEHYLGKDHSEDPKKNKAD----- 721
QY 707 -----EEPEETPAPEVPQVETEKEVAOLKEAEVLLAKVTDSL 746
```

DB 722 EETPREEKQSEKPSPKPTEEPSESEPESEPEVETEKVEKLEAREADLGKIODPII 781  
QY 747 KANAFETLAGLRNNLTQIMDNNSIMAEKILLALKGS 785  
DB 782 KSNAKETTLTGLKNLLFGTODNNNTIWAEEKILLALKES 820  
RESULT 5  
AAB12766  
ID AAB12766 standard: Protein: 821 AA.  
XX AAB12766;  
AC AAB12766;  
XX 21-NOV-2000 (first entry)  
DE Streptococcus pneumoniae strain SP64 BVH-11 protein antigen.  
XX  
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal.  
XX  
OS Streptococcus pneumoniae.  
XX WO200039299-A2.  
XX 06-JUL-2000.  
XX 20-DEC-1999; 99WO-CA01218.  
XX 23-DEC-1998; 98US-0113800.  
XX (BIOC-) BIOCHEM PHARMA INC.  
XX  
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
DR WPI; 2000-452397/39.  
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia -  
XX  
PS Disclosure: Fig 12; 106pp; English.  
XX  
XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens  
CC have bactericidal activity. The nucleic acids, encoding the protein  
CC antigens, may be used for the recombinant production of the proteins  
CC they encode. The protein antigens may then be used as vaccines for the  
CC prevention and treatment of Streptococcal infections in mammals  
CC (especially humans) which result in, e.g. meningitis, otitis media,  
CC bacteraemia and/or pneumonia. The present sequence represents a  
CC S. pneumoniae BVH-11 protein antigen, from the present invention.  
XX  
SQ Sequence 821 AA:  
Query Match 77.3%; Score 3218; DB 21; Length 821;  
Best Local Similarity 75.2%; Pred. No. 2.8e-229;  
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;  
QY 1 STELGILQARYKENVKRRSYIDGKATOKTENLTPEDESKREGINAEDQIVIKITDQGYVT 60  
DB 2 AVELGLHQATQYKENVKRRSYIDGKATOKTENLTPEDESKREGINAEDQIVIKITDQGYVT 61  
QY 61 SHGDHYHYNGVVPYDAITISELLMKDPYKXKDEDIVNEVGVYKVDGKYYVYLKDA 120  
DB 62 SHGDHYHYNGVVPYDAITISELLMKDPYKXKDEDIVNEVGVYKVDGKYYVYLKDA 121  
QY 121 AHADNVRTEKEELNRQKQSHREGSTPRNDGAVALARSGQRYTDDGYIFNASDIIDET 180  
DB 122 AHADNVRTEKEELNRQKQSHREGSTPRNDGAVALARSGQRYTDDGYIFNASDIIDET 181  
QY 181 GDAYIVPHGDHYHYIPKNELASSELAAEAFILSGRGNLSNSTRYRONSNDNTSRINWVPS 240

DB 182 GDAYIVPHGDHYHYIPKNELASSELAAEAFILSGRGNLSNSTRYRONSNDNTSRINWVPS 241  
QY 241 VSNPGTNTNTNNSNTNSQASOSNDIDSLKOLYKPLPSORHVESDGLVFPDAQTTSRT 300  
DB 242 VSNPGTNTNTNNSNTNSQASOSNDIDSLKOLYKPLPSORHVESDGLVFPDAQTTSRT 301  
QY 301 ARGVAVPHGDHYHYFTPYSSMSELEERIALIPLRYRSNMVPSDRPEOPSPOTPEPSPG 360  
DB 302 ARGVAVPHGNHNYHFIPYEDMSLEKRIARILPLRYRSNMVPSDRPEOPSPOTPEPSPS 361  
QY 361 POPAPNPK-IDSN---SSLVSQLVRKVGEGYVEEKGISRYFAADLSEYTKNLESKS 416  
DB 362 POPAPNPOPAPNSPIDEKLYKAVRVGVGVVEENGYSRYIPAKNLSAETAAGIDSKLA 421  
QY 417 KOESVHTLTAKKENVAPRDOEFYDKAYNLLEAKKALEFXNGRNSDFQALDKLERLND 476  
DB 422 KOESLSHKIGAKKTDLPSSDREFYKAYDLKRIHODLLDNKGROYDFALNLERLND 481  
QY 477 ESTNREKLVDDLLAFAPITTHPERLGKPNQIETYEDEVRIAOADKYTTSOGYTFDEHD 536  
DB 482 VSDKVKLVDDILAFAPIRHBERLGKPNQAQITTYDDDEIQVAKLAGKYTTEGDIYFDPD 541  
QY 537 IISDEGDAYVTPHMGSHWIGKDSLSDEKVAQAATKKEGILPPSPDADVYANPFGSA 596  
DB 542 IISDEGDAYVTPHMGSHWIKKDSLEAERAAQAAYAKKEGILPPSTDQDSGNTPEAKGA 601  
QY 597 AAIYNRVKGEKRLPLVRLPYVHEHYEVKNGNLIIPKDHYNIRKPAFDDHTKAPNGY 656  
DB 602 EAIYNRVKAKKAVPLDRMYNLQYIVKNGSLIIPHYDHNIRKPAFDDHTKAPNGY 661  
QY 657 TLEDLFATIKYVVEHPDERPHSDNGNASEHYLGKKHSDPNNKFKAD----- 706  
DB 662 TLEDLATYKYYVEHPDERPHSDNGNASEHYLGKKHSDPNNKFKAD----- 721  
QY 707 -----EPPVETPAEPVPOVETEKVPAOLKEAVYLLAKYTDSL 746.  
DB 722 EETPREEKQSEKPSPKPTEEPSESEPESEPEVETEKVEKLEAREADLGKIODPII 781  
QY 747 KANAFETLAGLRNNLTQIMDNNSIMAEKILLALKGS 785  
DB 782 KSNAKETTLTGLKNLLFGTODNNNTIWAEEKILLALKES 820  
RESULT 6  
AAU84026  
ID AAU84026 standard: Peptide: 821 AA.  
XX  
XX AAU84026;  
AC AAU84026;  
XX  
XX 08-MAY-2002 (first entry)  
DE Truncated variant of S. pneumoniae BVH-11, BVH-11M.  
XX  
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;  
KW pneumonia; streptococcal bacterial infection; mutant; mutein.  
OS Streptococcus pneumoniae.  
XX Synthetic.  
XX WO200198334-A2.  
XX  
XX 27-DEC-2001.  
XX  
XX 19-JUN-2001; 2001WO-CA00908.  
XX  
XX 20-JUN-2000; 2000US-212683P.  
XX  
XX (SHIR-) SHIRE BIOCHEM INC.  
XX  
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;  
DR WPI; 2002-122272/16.  
XX

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and  
PT epitope-bearing polypeptides, useful as vaccine components for treating  
PT or preventing streptococcal infections such as otitis media,  
PT meningitis, and bacteraemia

XX Example 1; Page -; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%  
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of  
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)  
CC comprising (I) is useful for therapeutic or prophylactic treatment of  
CC meningitis, otitis media, bacteraemia or pneumonia infection in an  
CC individual susceptible to these disorders. (II) is also useful for  
CC therapeutic or prophylactic treatment of any streptococcal bacterial  
CC infection (e.g., caused by Streptococcus pneumoniae, group A  
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such  
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or  
CC Streptococcus aureus) in an individual susceptible to the infection.  
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation  
CC techniques. The Streptococcus polypeptides are useful in a diagnostic  
CC test for S. pneumoniae infection. (III) is useful for designing DNA  
CC probes for use in detecting the presence of Streptococcus in a biological  
CC sample suspected of containing the bacteria. The DNA probes may also be  
CC used for detecting circulating S. pneumonia nucleic acid in a sample for  
CC diagnosing streptococcal infections. This sequence represents a truncate  
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,  
CC Note: This sequence does not appear in the specification but has  
CC been created according to information given in the invention.

XX Sequence 821 AA:

Query Match 77.3%; Score 3218; DB 23; Length 821;

Best Local Similarity 75.2%; Pred. No. 2.8e-229;

Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKENNRVSYIDGKATQKTENTLPDEVSKREGINAEOIVIKITDGGYVT 60  
DB 2 AYELGLHQGYTVKENNRVSYIDGKATQKTENTLPDEVSKREGINAEOIVIKITDGGYVT 61  
QY 61 SHGDHHYNGKPYDAIISEELMKDPNYKLDKEDIVNEKGGYIKYDKGYVYLKDA 120  
DB 62 SHGDHHYNGKPYDAIISEELMKDPNYQLKDSIDIVEIKGGYIKYNGKGYVYLKDA 121  
QY 121 AHDNVRTEKEIRKOEHSOHREGTPRNDGVALARSGRTTDDGYTFNANSDIIEPT 180  
DB 122 AHDNVRTEKEIRKOEHSOHREGTSANDGAVAFARSOGRTTDDGYTFNANSDIIEPT 181  
QY 181 GDAYIVPHGDHYHYIPKNELSASELAABEFLSGRNLNSRTYRONSNDNTSRFNWVPS 240  
DB 182 GDAYIVPHGDHYHYIPKNELSASELAABEFLSGRENLSNLRTYRONSNDNTPRTMWVPS 241  
QY 241 VSNPRTNTNTNSNTNSQASQNSNDISLKLQYKLPISORHVESDGLVFPDPAQTTSRT 300  
DB 242 VSNPRTNTNTNSNTNSQASQNSNDISLKLQYKLPISORHVESDGLIFPDPAQTTSRT 301  
QY 301 AGGVAAPHGDHYHYIPYOSMSELEERIALIPLRYRSNMWVPSPREOPSPOPTPPSPG 360  
DB 302 AGGVAAPHGNHHTFYEDOMSELEEKRIARIPLRYRSNMWVPSPREOPSPOPTPPSPS 361  
QY 361 POPAPMLK-IDSN--SLSVQLVRKVGEGYVEEKISRYVEPAKDLPSVTYKNELSKLS 416  
DB 362 POPAPMPGPAPNPIDEKLYKEAVRKGVGVEENGVSRYIPAKNLSAETAAGDISKLA 421  
QY 417 KOESVSHITLAKKENVAPRDQERYDKAYNLTFAHKAFLXNKGKNSDFOQLKLERLND 476  
DB 422 KOESLSHKLGAKKTDLPSSDREFYKNAYDLARIHODLLDNKGRQVDFALNLLERLND 481  
QY 477 ESNKKEKLVDDLALAPLTHPERLGPNSQTEYTEDVRIQLAKYTSQGYIDEHD 536  
DB 482 VSSDKVATDDILALAPLTHPERLGPNSQTEYTEDVRIQLAKYTSQGYIDEHD 541  
QY 537 IISDEGDVATVPHMGSHWIGKDSLDERKVAQAAYTKKGGILPPSPDADVANKANPTGDSA 596

DB 542 IISDEGDVATVPHMGSHWIGKDSLSEADRAAQAAYAKKGLTPSTDHQDSGNTEAKGA 601  
QY 597 AAIYNRKGEKRIPLRYLPMYWEHYEVKNGNLIIPKHCHYNNIKFAWDDHTTAPAKPY 656  
DB 602 EAIYNRVKAARKVPLDRMPYNIQYTEVKNGLIIPHYHYNHIFKFEWDEGLYEAPKGY 661  
QY 657 TLEDJFATIKYVEHDEPHSDNGNASEHYLKGKHSDEPNKFKAD----- 706  
DB 662 TLEDJLATYKVEHDEPHSDNGNASEHYLKGKHSDEPNKFKAD----- 721  
QY 707 -----EEPEETPAPEVPYVETEKVNAOLKEAEVLLAKYTDSSL 746  
DB 722 EETPREKQSEKPESEPKTEPEEESPESEPEQYETEKVEKILREAPDILGKIQDPPI 781  
QY 747 KANATTTAGLRNNLTQLQIMDNNSIHAEEKLLALLKGS 785  
DB 782 KSNAKETTLGLKNNLIFGTQDNNTIWAEEKLLALLKES 820

Result 7

ID AAB12716 standard; Protein; 840 AA.

AC AAB12716;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-11 protein antigen SEQ ID NO:4.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

DR N-PSDB; AAA65731.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia -

PS Claim 18; Fig 4; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens  
CC have bactericidal activity. The nucleic acids, encoding the protein  
CC antigens, may be used for the recombinant production of the proteins  
CC they encode. The protein antigens may then be used as vaccines for the  
CC prevention and treatment of Streptococcal infections in mammals  
CC (especially humans) which result in, e.g. meningitis, otitis media,  
CC bacteraemia and/or pneumonia. The present sequence represents the  
CC S. pneumoniae BVH-11 protein antigen.

SQ Sequence 840 AA:

Query Match 77.3%; Score 3218; DB 21; Length 840;

Best Local Similarity 75.2%; Pred. No. 2.9e-229;

Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKENNRVSYIDGKATQKTENTLPDEVSKREGINAEOIVIKITDGGYVT 60



```

Db      21 AVELGHOQTVKENNRVSIDGKQATQKTENLTPDEVSKREGINAEOIVIKITDOGYVT
QY      61 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKDDIYNEVGIVYKDGKYYVYLKDA 120
Db      81 SHGDHYHYNGKVPYDAIISELLMKDPNYQDKSDIYNEIKGVYIKNGKYYVYLKDA 140
QY      121 AHADNVRTKEEINRQKQESHQREGTTPRNDGAVALARSQGRYTTDDGYIFNADSIIEPT 180
Db      141 AHADNVRTKEEINRQKQESHQREGTSANDGAVAFARSQGRYTTDDGYIFNADSIIEPT 200
QY      181 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRQNSDNTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRQNSDNTSRTNWVPS 260
QY      241 VSNPGTTNTNTNNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAQITST 300
Db      261 VSNPGTTNTNTNNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAQITST 320
QY      301 ARGVAVPHGDHYHYIPYSQMSLEERIRIIPLRYSNMHWVDSRPEQSPQPTPEPSPG 360
Db      321 ARGVAVPHGNHNYHFIPYEQMSLEKRIARIIPLRYSNMHWVDSRPEQSPQPTPEPSPS 380
QY      361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYFEEKGISRYVFAFDLPSETVKNLESKUS 416
Db      381 POPAPNPQAPNPIDEXKLVKEAVRKVGCGYFEEENGVSRYIPAKNLSAETAGIDSKLA 440
QY      417 KOESVSHLTAKKENVAPRDOEFYKAVANLLTEAKKALFXNKGSRSDQALDKLERLND 476
Db      441 KOESLSHKLGAARKTDLPSDRFEYKKAADLARIHODLDNKGROVDEALDNLERLKD 500
QY      477 ESTNKEKLVDDLAFAPITPHERLGKPNQSEIYTEDVEYRIQALDKYTTSDGYTFDEHD 536
Db      501 VASDKVATLVDLAFAPITPHERLGKPNQAQITTYTDDIYQAKLAGKYTTEDEGYTFDPRD 560
QY      537 IISDEGDVATVPHMGSHWIGKDSLSDEKVAQAQAYTEKGLPPSPDADVANKPTGSA 596
Db      561 IISDEGDVATVPHMGSHWIGKDSLSDEKVAQAQAYTEKGLPPSTHODQSDGNTAKGA 620
QY      597 AAIYRVKGEKRIPLVRLPYVWEHYVEYVNGNLIIIPKHQYININIFAMWDDHITAPANGY 656
Db      621 EATYRVKAKKAVPLDRMPYNIQYTVVEYKNGSLIIPHDHYNINIKFEMFDEGLTEPKCY 680
QY      657 TLEDLFATIKYVVEHPDERPHSNDMGNASEHVLGKDKHSEDPNKNFKAD----- 706
Db      681 TLEDLATYKYVVEHPDERPHSDNGFGNASHYQNRKNQADPTNTEKPESEKPTQTEKPE 740
QY      707 -----EPPVEETPAPEVYPOVETEKVPAQLKEAEVLLAKYVTDSSL 746
Db      741 EETPREEKPOSEKPEPKPTTEPEEESPEESEPQVETEKVEKILREABDLGKIQDPIT 800
QY      747 KANATETLAGLNNLTLOIMONNSIMAEKLLALLKGS 785
Db      801 KSNAKETTLGLKNLLFGTQDNNTTMAEAKLLALLKES 839

RESULT
8 AAU75933
ID AAU75933 standard; Protein: 840 AA.
XX AAU75933;
AC AAU75933;
DT 08-MAY-2002 (first entry)
DE Streptococcus pneumoniae BVH-11 protein.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumoniae; streptococcal bacterial infection.
XX Streptococcus pneumoniae.
OS WO200198334-A2.
XX PN
XX

```

```

PD      27-DEC-2001.
XX      19-JUN-2001; 2001WO-CA00908.
PF      20-JUN-2000; 2000US-212683P.
PR      (SHIR-) SHIRE BIOCHEM INC.
PA      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI      WPI: 2002-122272/16.
DR      N-PSDB; ABK15103.
DR      WPI: 2002-122272/16.
XX      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT      epitope-bearing polypeptides, useful as vaccine components for treating
PT      or preventing streptococcal infections such as otitis media,
PR      meningitis, and bacteraemia.
XX      Example 1; Fig 7; 113pp; English.
PS      The invention describes an isolated polypeptide (I) with 70-90%
XX      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC      comprising (I) is useful for therapeutic or prophylactic treatment of
CC      meningitis, otitis media, bacteraemia or pneumonia infection in an
CC      individual susceptible to these disorders. (II) is also useful for
CC      therapeutic or prophylactic treatment of any streptococcal bacterial
CC      infection (e.g., caused by Streptococcus pneumoniae, group A
CC      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
CC      Streptococcus aureus) in an individual susceptible to the infection.
CC      A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC      techniques. The Streptococcus polypeptides are useful for diagnostic
CC      test for S. pneumoniae infection. (III) is useful for designing DNA
CC      probes for use in detecting the presence of Streptococcus in a biological
CC      sample suspected of containing the bacteria. The DNA probes may also be
CC      used for detecting circulating S. pneumonia nucleic acid in a sample for
CC      diagnosing streptococcal infections. This is the amino acid sequence of
CC      Streptococcus pneumoniae protein BVH-11, used to create the antigenic
CC      peptides described in the method of the invention.
SQ      Sequence 840 AA:
      77.3%; Score 3218; DB 23; Length 840;
Query Match 75.2%; Pred. No. 2,9e-229;
Best Local Similarity 65; Mismatches 104; Indels 34; Gaps 3;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY      1 SYELIGYQARTVKENNRVSIDGKQATQKTENLTPDEVSKREGINAEOIVIKITDOGYVT 60
Db      21 AVELGHOQTVKENNRVSIDGKQATQKTENLTPDEVSKREGINAEOIVIKITDOGYVT 80
QY      61 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKDDIYNEVGIVYKDGKYYVYLKDA 120
Db      81 SHGDHYHYNGKVPYDAIISELLMKDPNYQDKSDIYNEIKGVYIKNGKYYVYLKDA 140
QY      121 AHADNVRTKEEINRQKQESHQREGTTPRNDGAVALARSQGRYTTDDGYIFNADSIIEPT 180
Db      141 AHADNVRTKEEINRQKQESHQREGTSANDGAVAFARSQGRYTTDDGYIFNADSIIEPT 200
QY      181 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRQNSDNTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRQNSDNTSRTNWVPS 260
QY      241 VSNPGTTNTNTNNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAQITST 300
Db      261 VSNPGTTNTNTNNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAQITST 320
QY      301 ARGVAVPHGDHYHYIPYSQMSLEERIRIIPLRYSNMHWVDSRPEQSPQPTPEPSPG 360
Db      321 ARGVAVPHGNHNYHFIPYEQMSLEKRIARIIPLRYSNMHWVDSRPEQSPQPTPEPSPS 380
QY      361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYFEEKGISRYVFAFDLPSETVKNLESKUS 416

```

Db	361	POPAPNPQAPASNPIDDKLYKVAERKVGQGYVEENGVSRTYIPAPNLSAETAGIDSKLA	440
Oy	417	KOESVSHTLAKKENVAPRDQEFYDKAYNLLTEFAKALFXNKGNSDPQALDKLERLND	476
Db	441	KOESISHTLGAKKTDLPSSDREFNKAAYDLLARIHODLLDNGKRGVDFEALDNLLERLKD	500
Oy	477	ESTNKEKLYVDLLAFLAPITHPERLCKPNSQILETEDEVERIAQLADKYTTSQGYTFEDHD	536
Db	501	VSSDVKLYVDLILAEFLAPIRHPERLCKPNAQITTYDDELQVAKLAGKYTTEDEGYTFEDRD	560
Oy	537	IISDEGDAYPYPHAGSHHWIGKDSLSDKEKVAQAAYTKKGLPSPDADYKAPNTGDSA	596
Db	561	ITSDEGDAYPYPHHTFTHHWIKKOSLSABEAAAQAATAKCKGLPSTBODSGNTAEAGA	620
Oy	597	AAIYNRKGEKRIPLVRLPYWVEHVEVKNGLILPHKDHYHNKIFAWFDHTFYKAPNGY	656
Db	621	EAIYNRYAAKAKVPLDRLMPNLQATVEVKNGLIIPHYDHYHNKIFEMFDEGLYEAAPGY	680
Oy	657	TLEDLFAIKYVYHHPERHSNDGKGNASEHVLGKKHSDSEPNKNEKAD-----	706
Db	681	TLEDLTLATVKKYVEHPNERHSDNGFGMASDHYQRKNQOADTNOTEKSESEKPOTEKPE	740
Oy	707	-----EEPVEETPAPEVPOVETEKEVEAOKEAEVLLAKYTTDSSL	746
Db	741	EETPREKPOSEKESKRPTEPEPEESPESEPOVETKVEBEKLRBADLLGKIQDPIT	800
Oy	747	KANATETLAGLRNNLLQIINDNNSINAAEKKLLALLKGS	785
Db	801	KSNAKETLTGLKNNLLFGTODNNTIMAEAKLLALLKES	839
RESULT 9			
AA12745	ID	AA12745 standard; Protein; 690 AA.	
XX	AA12745:		
DT	23-NOV-2000	(first entry)	
XX	Streptococcus pneumoniae NEW16	protein antigen SEQ ID NO:79.	
DE	7		
KW	Streptococcus pneumoniae: BWH-3; BWH-11; BWH-28; antigen; vaccine;		
KM	prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;		
KW	otitis media; pneumonia; immunisation; bactericidal.		
XX	Streptococcus pneumoniae.		
OS			
PN	WO200039299-A2.		
XX	06-JUL-2000.		
PD			
XX	20-DEC-1999;	99WO-CA01218.	
PF			
XX	23-DEC-1998;	98US-0113800.	
PR			
PA	(BIOC-) BIOCHEM PHARMA INC.		
PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;		
DR	WPI: 2000-452397/39.		
XX	Streptococcal antigens useful for vaccinating against e.g. meningitis,		
PT	otitis media, bacteremia and/or pneumonia -		
XX			
PS	Claim 18; Fig 44; 106pp; English.		
XX	The present invention describes nucleic acids (I) encoding protein		
CC	antigens (II) from Streptococcus pneumoniae. The protein antigens		
CC	have bactericidal activity. The nucleic acids, encoding the protein		
CC	antigens, may be used for the recombinant production of the proteins		
CC	they encode. The protein antigens may then be used as vaccines for the		
CC	prevention and treatment of Streptococcal infections in mammals		
CC	(especially humans) which result in, e.g. meningitis, otitis media,		

CC	bacteraemia and/or pneumonia. The present sequence represents the
CC	S. pneumoniae NEW16 protein antigen.
XX	
XX	Sequence 690 AA;
Query Match	71.1%; Score 2961; DB 21; Length 690;
Best Local Similarity	80.7%; Pred. No. 2.3e-210;
Matches 556;	Conservative 54; Mismatches 75; Indels 4; Gaps 2.
QY	1 SYELGLQARTYKENVNRSYIDGKQATOKTENLTPEDESKREGINAEOIVIKITDQGYVT 60
DB	2 AVELGLHQOQTYKENVNRSYIDGKQATOKTENLTPEDESKREGINAEOIVIKITDQGYVT 61
QY	61 SHGDHYHYNGKVPYDAITISEELKADPNYKDKEDIVNEVKGQYVYKQDGYVYLKDA 120
DB	62 SHGDHYHYNGKVPYDAITISEELKADPNYKDKEDIVNEVKGQYVYKQDGYVYLKDA 121
QY	121 AAHDNRTEKEELNRQOEHSQHREGTSPNDCAVAVLARSGQVYTTDDGIIFNASDIETD 180
DB	122 AAHDNRTEKEELNRQOEHSQHREGTSDANDAVAVARSGQVYTTDDGIIFNASDIETD 181
QY	181 GDAYIPHGDHYIIPKNELSASELAAEAFLSGRGLNSRTYRONSNDNTSRTNWVS 240
DB	182 GDAYIPHGDHYIIPKNELSASELAAEAFLSGRGLNSLTTRYRONSNDNTPRTNWVS 241
QY	241 VSNPGTTNTSNTSNTSQAQSQNDISDLKQLYKPLSQRHVESDGLVEDPAQITSRT 300
DB	242 VSNPGTTNTSNTSNTSQAQSQNDISDLKQLYKPLSQRHVESDGLIFDPAQITSRT 301
QY	301 AGVAVPHGDHYHFIYQSMSLEERTIATITPLRYSNHWVDSRPEQSPQTPPEPSPG 360
DB	302 AGVAVPHGDHYHFIYEQMSLEERTIATITPLRYSNHWVDSRPEEPSPQTPPEPSPS 361
QY	361 POPAPYLRK-IDSN--SSLYSOLVLRKVGEGYFEEKGISRYFAKDPSETYKKNLESKLS 416
DB	362 POPAPYLRK-IDSN--SSLYSOLVLRKVGEGYFEEKGISRYFAKDPSETYKKNLESKLS 421
QY	417 KOESVSHITLAKENYAPRPODEFYDKAIVLLTEAHKALFYKNGRNSDFOALDKLERLD 476
DB	422 KOESLSHKLGAKKTDLPSSDRFENYKAYDLLRIHODLLDNKGROVDPEALNLLERLKD 481
QY	477 ESNRNEKVLVDLLAFLAPLTHPERLRGKPSQILEYDEVRINQOLADKYTSGYIFDEHD 536
DB	482 VSSDKKLVLDLILAFAPLRHPERLRGKPSQILEYDEVRINQOLADKYTSGYIFDPRD 541
QY	537 ITSDCEDAVYTPHMGSHWIGKDSLDSEKKEVAAQAATYKKEGILPSPDADYKANPTGDSA 596
DB	542 ITSDCEDAVYTPHMGSHWIGKDSLDSEKKEVAAQAATYKKEGILPSPDADYKANPTGDSA 601
QY	597 AAIYNEVKGEKRIPLVRLPYWEHTYEVKNGNLLIIPKHDIYHNIRKFAFDDHTYAPNCG 656
DB	602 EAIYNEVKKAARKVPLDRMPYNIQYTYEVKNGSLIIPHYDHYHNIRKFEWFDGELYEARQY 661
QY	657 TLEDLFAITIKYVEHDERPHSNDGNGNA 665
DB	662 TLEDLFAITIKYVEHDERPHSNDGNGNA 690
RESULT 10	
AA084093	
ID	AA084093 standard; Peptide: 690 AA.
XX	
AC	AA084093;
XX	
DT	08-MAY-2002 (first entry)
XX	
DE	Truncated variant of S. pneumoniae BVH-11, NEW16.
XX	
KM	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX	pneumonia; streptococcal bacterial infection; mutant; mutuin.
OS	Streptococcus pneumoniae.
XX	



CC particular, the polypeptides are useful for stimulating the immune  
 CC system and are effective to immunize or treat a mammalian subject  
 CC against Streptococcus pneumoniae infection or colonization.

xx Sequence 826 AA:

Query Match 67.5%; Score 2812; DB 21; Length 826;  
 Best Local Similarity 67.4%; Pred. No. 3.3e-199;  
 Matches 552; Conservative 81; Mismatches 138; Indels 48; Gaps 8;

```

QY 1 SYELGLYQARTV-KENNRYSYIDGKQATOKTENLPDEYSKREGINABOIVIKITDQGYV 59
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 21 SYELGRHQAGQVKKSNRYAIDGQAGKAENLPDEYSKREGINABOIVIKITDQGYV 80
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 TSHGDHYHYNYNGKVPYDAIISELLMKDPNYKLDKEDIYNEVKGQYVYKVDGKYVYLKD 119
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 81 TSHGDHYHYNYNGKVPYDAIISELLMKDPNYKLDKEDIYNEVKGQYVYKVDGKYVYLKD 140
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 AAHADNVRTKEEINROKQEHSHQREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIED 179
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 141 AAHADNIRTKKEIKRQKQSHSHHGGS--NDQAVVAARAQGYTTDDGYIFNASDIIED 198
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 180 TGDATYVPHGDHYHYIPKNELASASELAAPFLSGRGLMSRRTYRQNSDNTSTNNWP 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 TGDATYVPHGDHYHYIPKNELASASELAAPFLSGRGLMSRRTYRQNSDNTSTNNWP 247
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 240 SVSNPGTNTNTSNNSTNSQASQNSNDISLKLQYKLPQSORHVESDGLFDPQAQITSR 299
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 248 NPAQPRLSNNHMLVTPTTHQ--NQGENTISLLRELAKLSERHVESDGLFDPQAQITSR 306
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 300 TARGAIVPHGDHYHYIPYSQMSLEERIRIITPLRYRSNHWVPDSRPOSPQPTPEPSP 359
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 307 TARGAIVPHGDHYHYIPYSQMSLEERIRIITPLRYRSNHWVPDSRPOSPQPTPEPSP 366
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 360 GPOAPNLK-IDSN--SSLYSQLVKQVGEYVEEKGISRVKVAPELSEFYKNLESTL 415
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 SQQAPNPQAPASNPDELKVAEYKAGDGVYEEENGSKRTIPADLSAETPAAGDSTKL 426
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 416 SHOESVSHLTAKKENVARDQEFYDKAYNLLTEAKHAKLFXNKGNSDFQALDKLERLN 475
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 427 AKOESLSHKLAKKTDLPSSDREFYKAYDILLARIHQDLDDKGRQVDEALDNLLEKX 486
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 476 DSTNKEKIVDOLLAFIATTPERLGRKNSQIETTEDVYRQAQADKTTSDGYTFDEH 535
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 487 DVPSSDKVLDIIDLAFIARIRPERLGRKNAQIITYDDIIOYAKLAGKTYTDEGYTFDPR 546
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 536 DIISDEGDAYVTPHMGSHWIGKDSLDEKRYAAOAYTREKGIPLPPSPADYKANTQDS 595
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 547 DITSDEGDAYVTPHMTSHWIKKDSLSEAKRYAAOAYTAKKEGILTPPSTDHQDSGNTKAG 606
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 596 AAIYNNRYKGERKIRPLVRLPYVEHTVEYKNGNLIIPKRDYHNIRKFAWFDHTYKAPNG 655
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 607 ABAIYNNRYKAAKVPDLDRMPYNIQYTVYEVKNGSLIIPHYDHYNIRKFEWFDGLYEAAPKG 666
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 656 YLLEDELFAITIKYVHPERPHSNDGWNASDEHVLGKDHSDPKNKRADEP-----P 709
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 667 YLLEDELFAITIKYVHPERPHSNDGWNASDEHVLGKDHSDPKNKRADEP-----P 726
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 710 VEET-----PAEPVPOVETEKYFAOLKEAEVLLAKYTTDSSL 746
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 727 EETTREREKPOSEKPESEKPTPEEPSESEPEVETEKYFAOLKEAEVLLAKYTTDSSL 786
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 747 KANATETLAGLRNNLTLOIMDNNSIMAEAKDLALCKGS 785
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 787 KSNAKETTLGLKNNLLFTGQDNNTIMAEAKDLALCKGS 825
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 12  
 AAY81662  
 ID AAY81662 standard; Protein; 827 AA.

xx AC  
 AAY81662;

DT 24-MAY-2000 (first entry)

xx Streptococcus pneumoniae protein sequence ID311.

xx Streptococcus pneumoniae; vaccine; screening; protein antigen;

xx antibacterial; antiinflammatory; meningitis; infection; diagnosis;

xx Streptococcus pneumoniae.

xx MO200006737-A2.

xx 10-FEB-2000.

xx 27-JUL-1999; 99MO-GB02451.

xx 27-JUL-1998; 98GB-0016337.

xx 19-MAR-1999; 99OS-0125164.

xx (MICR-) MICROBIAL TECHNIQS LTD.

xx Gilbert CFG, Hansbro PM;

xx WPI; 2000-199300/17.

xx New Streptococcal protein, useful as a vaccine, for diagnosis of

xx Streptococcal diseases and for screening agents capable of antagonizing

xx or inhibiting expression of the protein

xx Claim 2; Page 99-100; 108pp; English.

xx AAY81501 to AAY81679 represent specifically claimed protein sequences

xx isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent

xx specifically claimed nucleotide sequences isolated from S. pneumoniae.

xx The sequences have antibacterial and antiinflammatory properties.

xx The protein sequences, and fragments of them, are useful as immunogens

xx and/or antigens. The nucleotide sequences can be used in vaccines and in

xx diagnostic assays. The proteins and nucleotides can be useful for the

xx detection and diagnosis of S. pneumoniae. The protein sequences are also

xx useful for screening an agent capable of antagonizing, inhibiting or

xx interfering with the function or expression of the proteins in which the

xx agent is useful for treatment or prophylaxis of S. pneumoniae infection

xx and meningitis. AA05591 to AA05614 represent primers used in the

xx exemplification of the present invention.

xx Sequence 827 AA;

xx Query Match 67.1%; Score 2795; DB 21; Length 827;

xx Best Local Similarity 67.0%; Pred. No. 5.9e-196;

xx Matches 549; Conservative 83; Mismatches 139; Indels 48; Gaps 8;

```

QY 1 SYELGLYQARTV-KENNRYSYIDGKQATOKTENLPDEYSKREGINABOIVIKITDQGYV 59
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 21 SYELGRHQAGQVKKSNRYAIDGQAGKAENLPDEYSKREGINABOIVIKITDQGYV 80
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 TSHGDHYHYNYNGKVPYDAIISELLMKDPNYKLDKEDIYNEVKGQYVYKVDGKYVYLKD 119
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 81 TSHGDHYHYNYNGKVPYDAIISELLMKDPNYKLDKEDIYNEVKGQYVYKVDGKYVYLKD 140
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 AAHADNVRTKEEINROKQEHSHQREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIED 179
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 141 AAHADNIRTKKEIKRQKQSHSHHGGS--GSGANDHAAVAAARAQGYTTDDGYIFNASDIIED 198
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 180 TGDATYVPHGDHYHYIPKNELASASELAAPFLSGRGLMSRRTYRQNSDNTSTNNWP 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 TGDATYVPHGDHYHYIPKNELASASELAAPFLSGRGLMSRRTYRQNSDNTSTNNWP 247
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 240 SVSNPGTNTNTSNNSTNSQASQNSNDISLKLQYKLPQSORHVESDGLFDPQAQITSR 299
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 248 NPAQPRLSNNHMLVTPTTHQ--NQGENTISLLRELAKLSERHVESDGLFDPQAQITSR 306
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 300 TARGAIVPHGDHYHYIPYSQMSLEERIRIITPLRYRSNHWVPDSRPOSPQPTPEPSP 359
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 307 TARGVAVPHGNHHPFIPYQMSLEKRIARIILPLRYRSNHWVDSRPQSPQSTPEPSP 366  
 QY 360 GPQAPNLK-IDSN---SSIVSOLVRKGVGEVEEKGISRVFKADLPSEVYKNLESL 415  
 Db 367 SQQPPNPQAPASNPIDEXKVAEYKVGDDGYVEENGVSRIIPAKDLSAETPAAGIDSL 426  
 QY 416 SKQESVSHLPFAKKENVAVRDQEFYDKAVNLLTEAHKALFXNKGNSDFQALDKLERLN 475  
 Db 427 AKQESLSHKLGAKKTDLPSSDREFYNKAYDLARIHODLLDKKGQVPEALDNLLERLK 486  
 QY 476 DESTKEKLVLDLAFAPITPERLGPNSQIETDEVRIAQIADKTTSDGIIFDEH 535  
 Db 487 DVPSQKAVLYDDIAFLAPIRHPERLGRNAQITTYDDEIOYAKLAGYTTEDGYIFDPR 546  
 QY 536 DIISDEGAYVTPPHGSHWIGKDSLDEKVAQAQYATREKGIILPSPADAVKANPTGDS 595  
 Db 547 DITSDEGAYVTPPHMTHSHWIKKDSLSEERAAQAQYAKKGLTPPSDHDQDSGTPEAKG 606  
 QY 596 AAATYNNRVKGEKRIPLVRLPYVWEHTVEVKNGLIIPKDHYNIRKFAWFDHTYKAPNG 655  
 Db 607 ABATYNNRVKAAKVPDLDRMPYNLQYTVETVKNGLIIPHYDHNIRKFEWFDGLTEAPRG 666  
 QY 656 YLLEDELATIKYVEHPDERPHSNDGKNASHEVYLGKDHSDPKNKNKADEE-----P 709  
 Db 667 YLLEDELATIKYVEHPDERPHSDNGFNASDHYORNNKGADTNQTEKPSSEKPOTEKP 726  
 QY 710 VEET-----PAPEVPOVETEKVPAOLKEAVLLAKYTDSL 746  
 Db 727 EETPREKPOSEKESKPTTEPESESEPESEPESEPESEPESEPESEPESEPESEPESE 786  
 QY 747 KANATETLAGLRNNLTLOIMDNNSIMAEAKLLALLKGS 785  
 Db 787 KSNAKETLTGLKNLNLFTGQDNNTIMAEAKLLALLKGS 825  
 RESULT 13  
 AAB01469  
 ID AAB01469 standard; Protein: 819 AA.  
 XX AAB01469;  
 AC AAB01469;  
 DT 20-OCT-2000 (first entry)  
 XX  
 DE Recombinant variant of Sp36 (Sp36B) of *S. pneumoniae*.  
 XX  
 KW Streptococcus pneumoniae; infection; vaccine; collid coll region;  
 KW histidine triad residue; Sp36; antibody; otitis media;  
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;  
 KW meningitis; lobar pneumonia.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 RN WO200037105-A2.  
 XX  
 PD 29-JUN-2000.  
 XX  
 PF 21-DEC-1999; 99WO-US30390.  
 XX  
 PR 21-DEC-1998; 98US-0113048.  
 XX  
 PA (MEDI-) MEDIMUNE INC.  
 XX  
 PI Johnson LS, Koenig S, Adamou JE;  
 XX WPI: 2000-452129/39.  
 DR N-PSDB; AAA47605.  
 XX  
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections  
 PT such as otitis media, nasopharyngeal and bronchial infections,  
 PT comprises Streptococcus pneumoniae proteins  
 XX  
 PS Claim 1; Page 65-69; 70pp; English.  
 XX

CC Although a number of proteins have been suggested as being involved  
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains  
 CC a need to identify polypeptides having epitopes in common from  
 CC various strains of *S. pneumoniae* in order to utilize such  
 CC polypeptides in vaccines to protect against a wide variety of  
 CC *S. pneumoniae*. New vaccine compositions are described which comprise a  
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino  
 CC acids in length that comprise at least one histidine triad residue  
 CC (HxxHxx) or a collid-coll region, or an antibody directed against  
 CC these features. The vaccine is useful in protecting against infection  
 CC by Streptococcus pneumoniae. The vaccine composition comprising  
 CC antibodies to is useful for passive immunization for treating  
 CC pneumococcal infections which includes otitis media, nasopharyngeal  
 CC and bronchial infections.  
 CC  
 XX  
 S0 Sequence 819 AA:  
 Query Match 67.0%; Score 2788.5; DB 21; Length 819;  
 Best Local Similarity 66.7%; Pred. No. 1.8e-197;  
 Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;  
 QY 1 SYELGLYQA-RIVKENNNVSYIDGKQATQKTENLTPDEYSKREGINAQIYKIIDGIV 59  
 Db 21 SYELGRYQAGQDKKSNRYAYIDGQAKAENLTPDEYSKREGINARQIYKIIDGIV 80  
 QY 60 TSHGDHYHYNNKVPYDAIISELLMKDPNFKDPEIVNEKGGYVIRKDGKYYVYLKD 119  
 Db 81 TSHGDHYHYNNKVPYDAIISELLMKDPNFKDPEIVNEKGGYVIRKDGKYYVYLKD 140  
 QY 120 AAHADNVTKKEIRINQKQESHQSHREGGTPRNDGAVALARASQGRYTTDGYIFNASDIED 179.  
 Db 141 AAHADNVTKKEIRINQKQESHQSHREGGTPRNDGAVALARASQGRYTTDGYIFNASDIED 197  
 QY 180 TGDATYVPHGCHYHYIPNNEISASFLAAGFLSGRGLNSRTRYRQNSDNTSRTKNVP 239  
 Db 198 TGDATYVPHGCHYHYIPNNEISASFLAAGFLSGRGLNSRTRYRQNSDNTSRTKNVP 246  
 QY 240 SVSNNGTNTNTNNSNSNSQASQNDIDSLKQYKPLPSORHESGLVDPQOTISR 299  
 Db 247 NPAQRLSENHLTYTPYHQ-NQGENISSLRELXAPLSRHHVESGLIFDPQOTISR 305  
 QY 300 TARGVAVPHGCHYHYIPYQMSLEKRIARIILPLRYRSNHWVDSRPQSPQSTPEPSP 359  
 Db 306 TARGVAVPHGNHHPFIPYQMSLEKRIARIILPLRYRSNHWVDSRPQSPQSTPEPSP 365  
 QY 360 GPQAPNLKIDSNSSVLQVAKVGEVYVEEKGISRVFKADLPSEVYKNLESLKQ 419  
 Db 366 SQQPPNPQAPASNPID--GRLVEAVKRVGDDGYVEENGVSRIIPAKDLSAETPAAGIDSLKQ 423  
 QY 420 SVSHTLTFAKKENVAVRDQEFYDKAVNLLTEAHKALFXNKGNSDFQALDKLERLNDEST 479  
 Db 424 SLSHKLTGKTKTDLPSSDREFYNKAYDLARIHODLLDKKGQVPEALDNLLERLKDVSS 483  
 QY 480 NKEKLVLDLAFAPITPERLGPNSQIETDEVRIAQIADKTTSDGIIFDEHDIIS 539  
 Db 484 DKVRLVEDILAFAPITPERLGPNSQIETDEVRIAQIADKTTSDGIIFDEHDIIS 543  
 QY 540 DEGDAYVTPPHGSHWIGKDSLDEKVAQAQYATREKGIILPSPADAVKANPTGSAAT 599  
 Db 544 DEGDAYVTPPHMTHSHWIKKDSLSEERAAQAQYAEKGLTPPSDHDQDSGTPEAKGAAEI 603  
 QY 600 YNNRVKGEKRIPLVRLPYVWEHTVEVKNGLIIPKDHYNIRKFAWFDHTYKAPGYTLE 659  
 Db 604 YNNRVKAAKVPDLDRMPYNLQYTVETVKNGLIIPHYDHNIRKFEWFDGLTEAPRGYTLE 663  
 QY 660 DLFAATIKYVEHPDERPHSNDGKNASHEVYLGKDHSDPKNKNKADEE-----PVEET 713  
 Db 664 DLATATIKYVEHPDERPHSNDGKNASHEVYLGKDHSDPKNKNKADEE-----PVEET 723  
 QY 714 -----PAPEVPOVETEKVPAOLKEAVLLAKYTDSLAKANA 750  
 Db 724 PREKPOSEKESKPTTEPESESEPESEPESEPESEPESEPESEPESEPESEPESE 783



PS Disclosure: Fig 12; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens  
CC have bactericidal activity. The nucleic acids, encoding the protein  
CC antigens, may be used for the recombinant production of the proteins  
CC they encode. The protein antigens may then be used as vaccines for the  
CC prevention and treatment of Streptococcal infections in mammals  
CC (especially humans) which result in, e.g. meningitis, otitis media,  
CC bacteraemia and/or pneumonia. The present sequence represents a  
CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.  
xx

Sequence 820 AA:

Query Match 66.6%; Score 2772; DB 21; Length 820;

Best Local Similarity 65.0%; Pred. No. 3e-196;

Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNVSYYIDGKQATQKTENTLPDEVSKREGINAEQIYIKITDQGYV 59  
DB 2 SYELGRHQAGQKKSNVSYYIDGQAGQKAKENLPPDEVSKREGINAEQIYIKITDQGYV 61  
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNKLKDEIVNEVKGGYVIRKVGKYYVYLKD 119  
DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNQLKDSDIYNEIKGGYVIRKVGKYYVYLKD 121  
QY 120 AAHADNVRTKEIRNROKQEHSGHREGTPRNDGAVALARSQGRYTTDDGYIFNADIIED 179  
DB 122 AAHADNITKTEIKRQKQESHNHGGS--NDQAVVAARAQGRYTTDDGYIFNADIIED 179  
QY 180 TGDAYIVPHGDHYHYIPKNELASELAAAEAFISGRGNLSRTRYRONSNTSRTNVP 239  
DB 180 TGDAYIVPHGDHYHYIPKNELASELAAAEAFYWG-----KGSRRPSSSSSYNA 228  
QY 240 SVSNPGTNTNTSNNSNTNSQASQNDISLKLQLYLPLSQRHVESDGLVDPQAITSR 299  
DB 229 NPAQRLSENNHLYTPYTHQ--NOGENISSILRELYAKPLSEHVESDGLIFDPAQITSR 287  
QY 300 TARGAIVPHGDHYHYIPYQSMELEERTARIIPLYRSNHNWPPDSRPQSPQSTPEPSP 359  
DB 288 TARGAIVPHGNHYHPIPEQSELEKRLARIIPLYRSNHNWPPDSRPQSPQSTPEPSP 347  
QY 360 GPQAPNLK-IDSN---SLVSQLVRKVGEGYVEEKGISRYVEAKDLPSSEYKNLESKL 415  
DB 348 SPQAPNPQAPASNPIDKLYKEARVKGQGVPEENGVSRTIPAKDLSAETAGIDSKL 407  
QY 416 SKQESVSHFTLAKKENVAPRQDEYDKAYNLTTEAKHALFYNNKGRNSDFQALDKLERLN 475  
DB 408 AKQESLSHKLGAKTDLPSDSREFYNAKAYDILLRIHODLNKGRQVDFEALDNLRLK 467  
QY 476 DESTNKEKLVDDLAFLAPITHPELGRPNQIETDEEVRIAQDLADKYTTSDGYIFDEH 535  
DB 468 DVPSDKVLYVDLILAFAPIRHPERLGRPNQIYTTDEIOVAKLAGKRYTTEDEGYIFDPR 527  
QY 536 DIISDEGDAYTTPHMGSHWIGKDSLSDEKVAQAQAYTKEGILPPSPDADYKANPFGDS 595  
DB 528 DITSDEGDAYTTPHMTSHWIKKDSLSEARAAQAYAKEGILPPSTDHODSGTTEAKG 587  
QY 596 AAATYNNRYKGEKRIPLVRLPYMVEHTVEVKNKGNLIIPHKDYHNKIFAMFDHDTYKAPNG 655  
DB 588 AEATYNNRYKAAKVPYLDPMYLNQYTVENKNGSLIIPHYDHYHNKIFEMFDEGLYEAPKG 647  
QY 656 YTLDELPAITIKYVYEHDPERPHSNDGNGNASEHYLG-----KDHSE----- 697  
DB 648 YTLDELPAITIKYVYEHDPERPHSNDGNGNASEHYLG-----KDHSE----- 697  
QY 698 -----DNNKNFKADEPEVEETPAPEVPEYETEKEVQAOLKEAEVLL 738  
DB 708 ESDKENHAGLNPSADNLYKRPSTDEETEBEABEDTTDEALIPQVENSYINAKIADAEPALL 767  
QY 739 AKAYDSSILKANATEETLAGLRNNLTQIMDNNSIMAEAEKLLALLKGSNPSV 790  
DB 768 EKVYDPSIRQANAMETLGLKSSLLGTGKDNNTISAQVDSLALLKESQAPPI 819

Search completed: November 12, 2002, 14:49:03  
Job time : 72 secs

**THIS PAGE BLANK (USPTO)**





```
|||||
Db 367 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVFAKDLPSBTVKLNESKLSKOS 426
Qy 421 VSHITLAKKENVAPROEYFDKAYNLTTEAHKALFNKGRNSDFQALDKLERLNDESTN 480
Db 427 VSHITLAKKENVAPROEYFDKAYNLTTEAHKALFNKGRNSDFQALDKLERLNDESTN 486
Qy 481 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIACLADKYTTSDGYIFDEHIIISD 540
Db 487 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIACLADKYTTSDGYIFDEHIIISD 546
Qy 541 EGDAYTTPHMGSHWIGKSLSDKEVAAQAYTKEGILPPSPDADYKANPTGDSAAIY 600
Db 547 EGDAYTTPHMGSHWIGKSLSDKEVAAQAYTKEGILPPSPDADYKANPTGDSAAIY 606
Qy 601 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDYHNKFAWFDHTYKAPNGYTTLED 660
Db 607 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDYHNKFAWFDHTYKAPNGYTTLED 666
Qy 661 LEATIKYVEHDEPDERPHSNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 720
Db 667 LEATIKYVEHDEPDERPHSNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 726
Qy 721 QVETEVKQALKEAEVLLAKYTDSSILKANATETLAGLRNNLTQIMDNNSIAAEAKLLA 780
Db 727 QVETEVKQALKEAEVLLAKYTDSSILKANATETLAGLRNNLTQIMDNNSIAAEAKLLA 786
Qy 781 LKGSNPSSVSKEKIN 796
Db 787 LKGSNPSSVSKEKIN 802
```

## RESULT 2

```
E98004
hypothetical protein phla [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; H
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; N
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-828 <KUP>
A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:g15458683; GSPDB:GN00174
C:Genetics:
A:Gene: phla
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 99.9%; Score 4159; DB 2; Length 828;
Best Local Similarity 99.7%; Pred. No. 3.2e-219;
Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 STEGLYOARTYKNNRVSIDGKATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT 60
Db 33 STEGLYOARTYKNNRVSIDGKATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT 92
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIYNEVGKGVYIKVGGYVYLKDA 120
Db 93 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIYNEVGKGVYIKVGGYVYLKDA 152
Qy 121 AHADNVRKEELNROKQSHREGSTPRNDGAVALANSQGRYTTDDGYIFNADSIIEDT 180
Db 153 AHADNVRKEELNROKQSHREGSTPRNDGAVALANSQGRYTTDDGYIFNADSIIEDT 212
Qy 181 GDAYIVPHGDHHTYIPKNELASSELAAEAFLISGRGNLSNSTYRROSDNSTRNWPVS 240
Db 213 GDAYIVPHGDHHTYIPKNELASSELAAEAFLISGRGNLSNSTYRROSDNSTRNWPVS 272
```

```
Qy 241 VSNPGTTNTNSNTNSOASOSNDISLLKQLYKPLSORHVESDGLVFPDAQITSTR 300
Db 273 VSNPGTTNTNSNTNSOASOSNDISLLKQLYKPLSORHVESDGLVFPDAQITSTR 332
Qy 301 ARGVAAPHGDHHTYIPYSOMSELEERIAIIPLRYSNMWVPDSRPEQSPQPTPEPSG 360
Db 333 ARGVAAPHGDHHTYIPYSOMSELEERIAIIPLRYSNMWVPDSRPEQSPQPTPEPSG 392
Qy 361 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVFAKDLPSBTVKLNESKLSKOS 420
Db 393 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVFAKDLPSBTVKLNESKLSKOS 452
Qy 421 VSHITLAKKENVAPROEYFDKAYNLTTEAHKALFNKGRNSDFQALDKLERLNDESTN 480
Db 453 VSHITLAKKENVAPROEYFDKAYNLTTEAHKALFNKGRNSDFQALDKLERLNDESTN 512
Qy 481 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIACLADKYTTSDGYIFDEHIIISD 540
Db 513 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIACLADKYTTSDGYIFDEHIIISD 572
Qy 541 EGDAYTTPHMGSHWIGKSLSDKEVAAQAYTKEGILPPSPDADYKANPTGDSAAIY 600
Db 573 EGDAYTTPHMGSHWIGKSLSDKEVAAQAYTKEGILPPSPDADYKANPTGDSAAIY 632
Qy 601 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDYHNKFAWFDHTYKAPNGYTTLED 660
Db 633 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDYHNKFAWFDHTYKAPNGYTTLED 692
Qy 661 LEATIKYVEHDEPDERPHSNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 720
Db 693 LEATIKYVEHDEPDERPHSNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 752
Qy 721 QVETEVKQALKEAEVLLAKYTDSSILKANATETLAGLRNNLTQIMDNNSIAAEAKLLA 780.
Db 753 QVETEVKQALKEAEVLLAKYTDSSILKANATETLAGLRNNLTQIMDNNSIAAEAKLLA 812
Qy 781 LKGSNPSSVSKEKIN 796
Db 813 LKGSNPSSVSKEKIN 828
```

## RESULT 3

```
D98004
histidine motif-containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUP>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:g15458682; GSPDB:GN00174
C:Genetics:
A:Gene: phla
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 77.4%; Score 3224; DB 2; Length 855;
Best Local Similarity 75.0%; Pred. No. 3.4e-168;
Matches 617; Conservative 69; Mismatches 99; Indels 38; Gaps 4;
```

```
Qy 1 STEGLYOARTYKNNRVSIDGKATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT 60
Db 32 STEGLYOARTYKNNRVSIDGKATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT 91
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIYNEVGKGVYIKVGGYVYLKDA 120
```

```
|||||
Db 92 SIGHDHYHNGVPPDAIIISELLMKDPNYQLKDEDIISEIKGGIVYKADGYYVYLKDA 151
QY 121 AAADNVRTKEEINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNADIIEDT 180
|||
Db 152 AAADNVRTKEEINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNADIIEDT 211
QY 181 GAYIVPHGDHYHYPKNELSASELAAPLISGRGNLSNSTRYRONSNTSRTNWVS 240
|||
Db 212 GAYIVPHGDHYHYPKNELSASELAAPLISGRGNLSNSTRYRONSNTSRTNWVS 271
QY 241 VSNPGTTNTNNTSNNTSQAOSNDISLLKOLYKPLSQRHVESDGLVDPQAOTTSRT 300
|||
Db 272 VSNPGTTNTNNTSNNTSQAOSNDISLLKOLYKPLSQRHVESDGLVDPQAOTTSRT 331
QY 301 AAGVAVPHGDHYHYPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQPTPEPSPG 360
|||
Db 332 AAGVAVPHGDHYHYPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQPTPEPSPG 391
QY 361 POPAPNPK-IDSN---SSLVSQLVRKVGEGYFEEKGISRYVFAKDLPSETVKNLESKLS 416
|||
Db 392 POPAPNPK-IDSN---SSLVSQLVRKVGEGYFEEKGISRYVFAKDLPSETVKNLESKLS 451
QY 417 KOESVSHLTAKKENAVAPRDQEFYDKAYNLLEAKKALFXNKGKRSDFQALDKLERLND 476
|||
Db 452 KOESVSHLTAKKENAVAPRDQEFYDKAYNLLEAKKALFXNKGKRSDFQALDKLERLND 511
QY 477 ESTNKEKLVDDLAFAPITPHERLGKPNQSEIETEDEVRIQOLADKYTSGYIFDEHD 536
|||
Db 512 VSSDKVYKLVDDLAFAPITPHERLGKPNQSEIETEDEVRIQOLADKYTSGYIFDEHD 571
QY 537 ITSDGDVAVPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSA 596
|||
Db 572 ITSDGDVAVPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSA 631
QY 597 AATYRVGKERIPVRLPYWVEHTVEYKNGLLIPHKDHYNINIFAFMDHTTYAPNXY 656
|||
Db 632 AATYRVGKERIPVRLPYWVEHTVEYKNGLLIPHKDHYNINIFAFMDHTTYAPNXY 691
QY 657 TLEDFATIKYVEHPRPHSDNGMNASEHVLGKDHSEDPNKFRADEE-----PV 710
|||
Db 692 SLEDLATYKYYVEHPRPHSDNGMNASEHVLGKDHSEDPNKFRADEE-----PV 751
QY 711 EET-----PAPEVPQVETEKVEAQLKEAEVLLAVT 742
|||
Db 752 EETPREKROSEKPEPPTPEPESESESESESESESESESESESESESESESESESE 811
QY 743 DSSLKANETETLAGRNLLTQIMDNNSIMAEKLLALLKGS 785
Db 812 NPIISNAKETILGLKNMLLFGTDNNNTIMAEKLLALLKGS 854

RESULT 4
895136
conserved domain protein Sp1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B95136
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap-
ple, Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:914972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1174
```

```
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 67.2%; Score 2797.5; DB 2; Length 819;
Best local similarity 67.0%; Pred. No. 5.9e-145;
Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;

QY 1 STELGLYQA-RIVKENNRVSYIDGKQATQKTENLAPDEVSKREGINAEDIVIKITDQGYV 59
|||
Db 21 SYELGRYQAGODKESNRVAYIDGQAKAENLTPDEVSKREGINAEDIVIKITDQGYV 80
QY 60 TSHGDHYHYNKVPDAIIISELLMKDPNYQLKDEDIISEIKGGIVYKADGYYVYLKDA 119
|||
Db 81 TSHGDHYHYNKVPDAIIISELLMKDPNYQLKDEDIISEIKGGIVYKADGYYVYLKDA 140
QY 120 AAADNVRTKEEINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNADIIED 179
|||
Db 141 AAADNVRTKEEINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNADIIED 197
QY 180 TGDVAVPHGDHYHYPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQPTPEPSPG 239
|||
Db 198 TGDVAVPHGDHYHYPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQPTPEPSPG 246
QY 240 SVSNPGTTNTNNTSNNTSQAOSNDISLLKOLYKPLSQRHVESDGLVDPQAOTTSRT 299
|||
Db 247 NPAQRLSENHMLTYTPYHQ--NOGENISLRLRELYAKPLSERHVESDGLVDPQAOTTSR 305
QY 300 TARGVAVPHGDHYHYPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQPTPEPSPG 359
|||
Db 306 TARGVAVPHGNHYHYPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQPTPEPSPG 365
QY 360 GPQAPNPKIDSNSSLVSQLVRKVGEGYFEEKGISRYVFAKDLPSETVKNLESKLSKOE 419
|||
Db 366 SPQAPNPNID--EKLVKAVKVGEGYFEEKGISRYVFAKDLPSETVKNLESKLSKOE 423
QY 420 SVSHLTAKKENAVAPRDQEFYDKAYNLLEAKKALFXNKGKRSDFQALDKLERLND 479
|||
Db 424 SVSHLTAKKENAVAPRDQEFYDKAYNLLEAKKALFXNKGKRSDFQALDKLERLND 483
QY 480 NKEKLVDDLAFAPITPHERLGKPNQSEIETEDEVRIQOLADKYTSGYIFDEHD 539
|||
Db 484 DKVYLEDLATLAPITPHERLGKPNQSEIETEDEVRIQOLADKYTSGYIFDEHD 543
QY 540 DEGDVAVPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSA 599
|||
Db 544 DEGDVAVPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSA 603
QY 600 YNRVAGKERIPVRLPYWVEHTVEYKNGLLIPHKDHYNINIFAFMDHTTYAPNXY 659
|||
Db 604 YNRVAGKERIPVRLPYWVEHTVEYKNGLLIPHKDHYNINIFAFMDHTTYAPNXY 663
QY 660 DLFATIKYVEHPRPHSDNGMNASEHVLGKDHSEDPNKFRADEE-----PV 713
|||
Db 664 DLFATIKYVEHPRPHSDNGMNASEHVLGKDHSEDPNKFRADEE-----PV 723
QY 714 -----PAPEVPQVETEKVEAQLKEAEVLLAVTSSIKANA 750
|||
Db 724 PREKROSEKPEPPTPEPESESESESESESESESESESESESESESESESESE 783
QY 751 TETLAGLNNLLTQIMDNNSIMAEKLLALLKGS 785
Db 784 KETTLGLKNMLLFGTDNNNTIMAEKLLALLKGS 818

RESULT 5
695115
conserved hypothetical protein Sp1003 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap-
ple, Science 293, 498-506, 2001
```



Db 666 YSLEDLATVYKYYEHPNERPHSDNGFCMASDHVQRNKGADTNOTKEPNEDEKPOTEKP 725  
Oy 702 -----NEKAD-----EPPVEPTAEPPEPOVE 723  
Db 726 EEDKEHDEVSPTPESEKENEHGLNPSADNLTKPSTDTETEETDEADDTDEAITPOVE 785  
Oy 724 TEKEVAOLKEAEVLLAKVYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAAEKLLALLK 783  
Db 786 HSVIAKIAEAEALLETXTDSSIRQNAVETLTGLKSSLLTGTQKNTTISAIEVDSLALLK 845  
Oy 784 GSNPSSV 790  
Db 846 ESQPTPI 852

## RESULT 7

conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain TIGR  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: H95115  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: H95115  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1039 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP1004

Query Match 29.9%; Score 1246; DB 2; Length 1039;  
Best Local Similarity 40.9%; Pred. No. 3.6e-60;  
Matches 296; Conservative 97; Mismatches 180; Indels 150; Gaps 18;

Oy 1 SYELGLQARTVKKENRNVSYIDGKQATQKTENLPDEYVKREGINAEQIVIKITDQGYT 60  
Db 22 AYALNQRHSQENKNNRVSVDGSSQKSENLPDQVSKREGIOAEQIVIKITDQGYT 81  
Oy 61 SHGHYHYNNKVPYDAIISPELLMKDPNYKTKDEDIVNEVKGQVYIKVDGKYVYVYLKA 120  
Db 82 SHGHYHYNNKVPYDAIISPELLMKDPNYKTKDEDIVNEVKGQVYIKVDGKYVYVYLKA 141  
Oy 121 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 180  
Db 142 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 197  
Oy 181 GDATVPRGHGHIYIPKNELSASELAAAEAFLSGRGNLSNRTYTRONSDNTSRTNWPS 240  
Db 198 GNAIVTPRGHGHYIPKSDLSASELAAAKAHLAGKNMOPOLSYSTFASDN----- 248  
Oy 241 VSNRGTNTNTSNNSTNSQASQSDNDISLLKOLYKLPLOSRIHESQGLVDPDAQITRT 300  
Db 249 -----NTQSVAKGSTSKPANKSENLOSLEKELYDSPAORYSESDGLVDPDAKIIISRT 301  
Oy 301 ARGVAVPHGHIYIPYQSOMSELEERIALIPLRYRSNHWPPDSRPEQSPQPPPEPSPG 360  
Db 302 PNGVAIPHGHIYIPYKSLALEEKIARMP----- 333  
Oy 361 POPAPNLKIDNSSLSVLQVAKVGEVGFEEKGISRYVFAKDLPESEYVKNIESKLSQES 420  
Db 334 -----ISGTGSTVSNNAK-----PNEVYSSLGSLSSNPS 363  
Oy 421 VSHTLTAKKENVAPRDOEFYKAVNLLTEAHKALEFXNKGGRNSDQADLKLRLINDESTN 480  
Db 364 -----LITSKELSSASDGIITNPK-DIVEETATATYIVRHG--DHFHYIPK-----SNOIG 410

Oy 481 KEKIVDLLAEFLAPITPHERBLGKPNQSIETYEDEVARIAQLADKYTSDGIYFDEHDIISD 540  
Db 411 OPTLPNNSLATPSP-SLPIINQTSHEKHE-----EGCYGFDAKRIITAE 452  
Oy 541 EGDAYVTPHMGSHWIGKDSLSDEKQVAAQAYTEKGIPLPSPDADKAPNPTGDSAAAY 600  
Db 453 DESGFVMSHGHNHYFFKDKLTDEQIKAAOKHLEB-----VTSNNGDLDSLSH 501  
Oy 601 NR-----VKKEKRIPLVRLPYW--EHTTEVKNGNLLI-PKQKHYNHAKRAM 644  
Db 502 EODYPSNAKEMKDLQKTEEKIAGIMQYGVKRESIVYVKKERNALITYPHGDHHAADP--- 558  
Oy 645 FDDHTYKAPNGYTLEDDFATIKYYEHPDERPHSNDG-----WGNASEHYLKKDHSDDPN 700  
Db 559 IDEH---KPVGIG---HSHSNYELFKPEBGVAKKEGKKYVTGBELTNVNLKNSFTNN 611  
Oy 701 KNF 703  
Db 612 QNF 614

## RESULT 8

hypothetical protein pHEC [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: D97985  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: D97985  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1039 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174  
C:Genetics:  
A:Gene: pHEC

Query Match 29.8%; Score 1243; DB 2; Length 1039;  
Best Local Similarity 41.2%; Pred. No. 5.3e-60;  
Matches 298; Conservative 94; Mismatches 180; Indels 152; Gaps 20;

Oy 1 SYELGLQARTVKKENRNVSYIDGKQATQKTENLPDEYVKREGINAEQIVIKITDQGYT 60  
Db 22 AYALNQRHSQENKNNRVSVDGSSQKSENLPDQVSKREGIOAEQIVIKITDQGYT 81  
Oy 61 SHGHYHYNNKVPYDAIISPELLMKDPNYKTKDEDIVNEVKGQVYIKVDGKYVYVYLKA 120  
Db 82 SHGHYHYNNKVPYDAIISPELLMKDPNYKTKDEDIVNEVKGQVYIKVDGKYVYVYLKA 141  
Oy 121 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 180  
Db 142 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 197  
Oy 181 GDATVPRGHGHIYIPKNELSASELAAAEAFLSGRGNLSNRTYTRONSDNTSRTNWPS 240  
Db 198 GNAIVTPRGHGHYIPKSDLSASELAAAKAHLAGKNMOPOLSYSTFASDN----- 248  
Oy 241 VSNRGTNTNTSNNSTNSQASQSDNDISLLKOLYKLPLOSRIHESQGLVDPDAQITRT 300  
Db 249 -----NTQSVAKGSTSKPANKSENLOSLEKELYDSPAORYSESDGLVDPDAKIIISRT 301  
Oy 301 ARGVAVPHGHIYIPYQSOMSELEERIALIPLRYRSNHWPPDSRPEQSPQPPPEPSPG 360  
Db 302 PNGVAIPHGHIYIPYKSLALEEKIARMP----- 333  
Oy 361 POPAPNLKIDNSSLSVLQVAKVGEVGFEEKGISRYVFAKDLPESEYVKNIESKLSQES 420  
Db 334 -----ISGTGSTVSNNAK-----PNEVYSSLGSLSSNPS 363

```

OY 421 VSHITAKKENVAPRDOEYDKAYMLLTENAKALXXNKRNSDFOLDKLERLNDSEIN 480
Db 364 ----LTTSEKLSASDGYTFNPK-DIVETATAYIVRHG--DHFYIPK-----SNOIG 410
OY 481 KEKVLDDLLAFIAPRTIHPERLQKP-NSQIEYFEDEVRIAGLADKXTTSDGYTFDEHDIIS 539
Db 411 QPTLLNNNSLAPSP-----SLPINPGISHKEHKE-----EDGYGFEDANRIIA 451
OY 540 DEGDAYVTPIRMHSHMWIGKDSLDEKEVAQAQVTEKGIPLSPDADYKANPTG-DSAA 598
Db 452 EDESGFIMSHGNHNHYFFKKDLTEQIKAQCKHLEE-----YKTSNGJDSLSS 500
OY 599 IYNRYKQ-----EKRIPLVRLPYWV-EHTVEVKNGLII-PIKHDIYNIKFA 643
Db 501 HQDYPGNAKEMKDLCKIIEKTIAGIMQGYVKRRSIVYNNKKNAIIPRQHHNADP-- 558
OY 644 WEDDHTTYAPNGYTLIEDLFATIKIYVENHDERPHSNDG---WGNASEHYLAKDHSDEP 699
Db 559 -IDEH---KPGVIG---HSHSNYELFPREEGVAKKGVYVTGEELTNVYMLLNSTFN 610
OY 700 NKNF 703
Db 611 NQNF 614

```

**RESULT 9**

T46758  
hypothetical g2.4k protein - Streptococcus agalactiae  
C:Species: Streptococcus agalactiae  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000  
R:Accession: T46758  
R:Spillerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heynenmann, J.; Schnitzler, N.; Lueg  
Infect. Immun. 67, 871-878, 1999  
A>Title: Lmb, a protein with similarities to the UraI adhesin family, mediates attachment  
A:Reference number: Z24091; MUID:99115568; PMID:9916102  
A:Accession: T46758  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-822 <SPe>  
A:Cross-references: EMBL:AFO62533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624  
A:Experimental source: strain R268  
C:Superfamily: Streptococcus agalactiae hypothetical g2.4k protein

Query Match            22.4% ; Score 934; DB 2; Length 822;  
Best Local Similarity   29.4%; Pred. No. 2,7e+43;  
Matches   271; Conservative   99; Mismatches   233; Indels   318; Gaps   32;

OY         1 SYELLLYOARVYKNNRRSYID--GKOATOKTEMLTLDVEYSKRGINAEQIVIKITDQG 57  
            |||:|||::|||:||||| | | | | | :|||:|||||:|||||  
DB       22 STQLCKHHMGGLATKNQALYIDDSSKGKVAKPRT-NKTMDQISAERGISAEOIVAVKITDGG 80  
  
OY       58 VYTSHGDHYHYYNGKVPAIDAISEELLAKDPNYKLAKDEDIVENVGGVYIKVDKKYYYVL 117  
            |||:|||:|||||:|||||:||||| | | | | | :|||:|||||:|||||  
DB       81 YVTSHGDHYHFNGKVYPDAIDISELLMTDPNYHKQSDVINELIDGYVIKVNNGYYYYL 140  
  
OY       118 KDAADADVNRKEELNROKHSGH-REGS-----TPRDNDAAVALARSGRITYTDGCT 169  
            |||:|||:|||||:|||||:||||| | | | | | :|||:|||||:|||||  
DB       141 KPGRSRKKNIIFTKQQIAEYAVAGKTGEAKERGLAQVAHLSEEVYAANEAARKRGRIITYDDIGY 200  
  
OY       170 IFNSADIIEDTGDAYIVPHGDHYHTIPKNELSASELAFAAFSL---GRGNLSNSTRYRR 226  
            |||:|||:|||||:|||||:||||| | | | | | :|||:|||||:|||||  
DB       201 IFSPTIIDDLGDAILVHGHNHYHTIPPKDLSPSLELAQAAYWSQKQGRG-AARSDVRP 258  
  
OY       227 QNSDNTSRTNMVPVSYS-NEGTTNTFTSN-----NSNTNSQASQSND----IDSLLK 272  
            |||:|||:|||||:|||||:||||| | | | | | :|||:|||||:|||||  
DB       259 TPAPGRARRAP-LPDVTPBPGOGHQPDNGCYHPARPBPNDASQNKHQREDFGKGTPEILLD 317  
  
OY       273 OLYXPLFLSORHESGGLFEDPAOLISRTARGVAIPHGHGYHPIFYSQMSELFEARTARIIP 332  
            |||:|||:|||||:|||||:||||| | | | | | :|||:|||||:|||||  
DB       318 OLHRIDLKYTRHEDEGLTFEPLOYIKSNAGLVVPHGDHYTHITPSOLFPLEMEJAD-- 374  
  
OY       333 LYRSNHMVWPPSRPQPPOPTPEPSPGPOPAPNLKIDSNSGLYSQLRKVGEGTYEEFK 392

[illegible][illegible]



```

Db 784 QNSQSEVEEIPVNEDEFEEDVHTQDLQDHLKTVDPVEIYEVEEIPSELHNEVAHPVEIETE 843
Oy 469 KLERLNDSTNKRKELVDLLAFLA-PIHPERLGRPNQSIETDEE-RIQLOADKYTT 526
Db 844 EYFPEPNQNNQNEQOINEDKSAHIOHEIYEVEEILPEDKNEKEVEHEIYEVEEILPEDKN 903
Oy 527 SDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDKREKVAQAAYTKRGILPPSPDAD 586
Db 904 EKG-----QHEIYEVEE---ELPDE-----DDKNEKEVEHEIYEVEEILP--EDKN 942
Oy 587 VKANPTGDSAAAIYNRVKGEKRIPLVRLPYMEH-IVFKNGNGLIIPHDHYNHKKFAMF 645
Db 943 EKGQHEIYEVEEILPEDKNEK-----VEHEIYEVEE---ILPEDKNEKG----- 983
Oy 646 DHTYKAPNGYTLDELFAITIKYVEHPDE-RPHSNDGMCNAS-----EHVLGKDHSE--- 697
Db 984 -QHEIYEVEEILPEDKNEKQOHEIYEVEEILPEDKNEKQOHEIYEVEEILPEDKNEKQ 1042
Oy 698 -----DPKNFKADEPEVEETPAPEVPOVETKEVEAQLKEAVLLAKVYDSSLK 747
Db 1043 HEIYEVEEILPEDKNEKQOHEIYEVEE---VEEILPEDKNEKQOHEIYEVEEILPEIYE--- 1095
Oy 748 NATETLALRNRLTQIMDNNSINAEKLLALLKGSNPSVSKKI 795
Db 1096 --IEEVPQTNNNENI-----TIPEEK-----NEFSVEEKAI 1128

RESULT 13
D90011
Fntb protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D90011
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D90011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2481 <KUR>
A:Cross-references: GB:BA000018; PID:g13701961; PIDN:BA043253.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fntB(mrp)

Query Match 4.1%; Score 172; DB 2; Length 2481;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 169; Conservative 140; Mismatches 353; Indels 210; Gaps 38;

Oy 8 OARTYKENRVSYIDGKQATOKT-----ENLTPDEVSKREGINAEQIYIKITDGYTS 61
Db 969 QAATTTKSAPKAEIAOKASEKRTALEAMNDSTTEEOAKKDY-----DQAVVTA 1018
Oy 62 HSD-----HYHYNGKVPYDAIISEELMKDPNFKLKDEDIVNEKGYVIRKDGKYYV 115
Db 1019 NADIDWATNTDYDAKAKTNEATIA--AITPDANVPRAKQAIAD-----KVQAO--- 1066
Oy 116 YLKDAHADNVKTKKEINQKQEHSHRGSGTFRNDGAVALAR----- 159
Db 1067 --ETIADANGSTTEKEKAQOVOTEKTA---DAAIDAASHNVEVEAANAETAKIE 1120
Oy 160 --QGRYTDGTFIENASDIIEDTGDAIYIPHGDNHYIRKNELSASELAAAEFLSGRN 217
Db 1121 AIQPATTTKD---NAKQAIATKANERKTA-----IAOTQITAEIAAANA-----D 1164
Oy 218 LSNRTYRQNSDNTSRTIMWVPSVSNPTGTTNTNTSNNSTNSQASQASNDISLKLQLYKL 277
Db 1165 VMAATYQANSNIEAANSQNDVQAKTGTSTID-QVTPYNNKKATARNITAILNN--KL 1221
Oy 278 PLQSHVESDGLVFDPAQITSTRAGVAVPHGDHYHIFIPYQMSLEERIAIRIPLRYS 337

```

```

Db 1222 QRIQATPPATDEKQAAAEANTENGA-----NOAISATTTNAQVDEKANA 1269
Oy 338 NIWVDSRPEQSPQPTPEPPSGQPA-----PNLKIDNSSLVQYRKVEG--- 386
Db 1270 EAIINAVTPKVVYKQAAKDEIDQLOATQTNVINDQNKATNEKEKAIDQLOATAVTDKNN 1329
Oy 387 --YVEEKGISRYFAKDLPSETVNLDS---KLSQESVSHTLAKKENVAPRQDEFY 440
Db 1330 ITAATDDKNGVDI---AKQAGKNSIGSTQPATAVKSNAKNEVDQAVTTQNOAT----- 1378
Oy 441 DKAIVLLIEAHKALFYXKGRNSDFQALDKLER-LNDSTN-----KEKYDDLAFIAP 494
Db 1379 DMTTGATTEE-----KNAKADLVLRKAKKAYODIINAQTTDVTQIKQAAVDIGITAD 1433
Oy 495 ITHPERLGPNQSIETDEEVRIQLOADKYTSDGYIFDEHDIISDEGDAYVTPHMGSH 554
Db 1434 TTIKD-VAKDELATYANQKALIAOTADAT-----EKKQANOQVDAQLT--QGNON 1483
Oy 555 WIGKDSLSDREKVAQAAYTKRGILPPSPDADYKANPTGDSAAAIYNRVKGEKRIPLVRL 614
Db 1484 IENASIDVNTAKDNAL---QALDPIQASTDVKTNARAEILTEMQNKI-----TEI 1532
Oy 615 PYMEHTVEYKNGNLIIP-----HKDHYHKKFAMF--DHTYKAPNGYTLDELFAITIKY 668
Db 1533 LNNNETTEEEK-GNDIGPRAAYEGGLNNINATTTGDTTAKDPAVOKVQDLHA----- 1586
Oy 669 VEHDPERPHSNDGMCNASEHVLGKKRDSHPNKNKADEE-PVEETP--AEPEVPOVETE 725
Db 1587 --NPYKKP-----AGKKELDQ-----AAQKKTQIEQTPNASQOEINDAQOE 1626
Oy 726 KYEAOIKEA-----EVLLAKVYDSSLKANAETPLAGIRNNLTIQIMD----- 767
Db 1627 -VDTEINQAKTAVNDOSSTNEYVDNAVEKGAIKINAVKTFSEKKDALKIEDAVNAKYNE 1685
Oy 768 -----NNSINAEKLLALLKGSNPSVSK 792
Db 1686 ADNSNASTSSEIAEAKQKALIELKQTAQDQNVNQ 1717

RESULT 14
S56271
Hypothetical protein YF016c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 19-Apr-2002
C:Accession: S56271
R:Murakami, Y.; Naitou, M.; Hagihara, H.; Shibata, T.; Ozawa, M.; Sasamura, S. I.; Sas
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces
A:Reference number: S56186
A:Accession: S56271
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1233 <MUR>
A:Cross-references: EMBL:D50617; NID:9836685; PIDN:BA09255.1; PID:d1009896; PID:9836
C:Genetics:
A:Map position: 6R

Query Match 4.1%; Score 171; DB 2; Length 1233;
Best Local Similarity 19.8%; Pred. No. 0.21;
Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;

Oy 12 VKENNRVSYIDGKQATOKTENTLPDEVSKREGINAEQIYI-----KITD--QGYTS 61
Db 151 IKETSTNNVABG-----TENVP--IKESTGIEVGNSPITPRKKKKKKTTNRGRSS 202
Oy 62 HGDNHYHYNGKVPYDAIIT--SEELMKDPNFKLKDEDI-----VNEKGYGVIVKD 110
Db 203 NPADTTDLKSQSTLDSILVIGIEVLEQDGS---KNEDIKVINIVQDEPVPNVEKMDIRTRNE 259
Oy 111 GKYIYLLKDAHADNV-----RTKEINRQKQEHSHRGSG--TPRNDGAVALARSQ--- 160

```



```

Db 260 SSDKTFIDIDPNKNDVETSSKSENNINEEKAHTLPRRENEILNVGNNASEFKHOLEP 319
Qy 161 -GRYTTDDGYIFENASDIETPDGDAYIVPHGDHYHYIPKNETLSASEFLAAEAFLSGRMIS 219
Db 320 HGLEAGDE-----NGQASIKDYSESELTKNGFNFKENESKHLKAGKQOTE---SPRDIS 372
Qy 220 NSRTYRONSQNTSTNNVPSVSNP-----GTTNTNTSNNSNTNSAQSQSDNI----- 267
Db 373 PSVLAKNQKEITEIGEDHVFQKDEKCEKRELNVNHNHNFNAAGSSITIPPETE 432
Qy 268 -----DILLKQLYKPLSORHY-----ESGLVFDPAQ---ITSR 299
Db 433 RETYDETMGPTKRISDEKMLQHGNTDISEVEKEEEEEEEENSTFSKYKKENVTGE 492
Qy 300 -----TARGVAVPHGDHYHFIPYSOMSELEBRIARIIPLRYSN----- 338
Db 493 QEAVRNNNEVSGTEESTSKGEIIMGD-----EKQSEAGRK-SSTIEIEGSAASATIS 544
Qy 339 --HWVPDSRPQSPQPPPEPSGPQAPANLKIDSNSSLVSQLYARKVGEGVFEKGIISR 396
Db 545 KDNVLDEDEAPQENKPTVEYGEIDIPDAPRD-----VEIYAVAEKNIIPEDLEVAK 599
Qy 397 -----YVAKDLPESTVYNLESKLSKQSVSHITLAKENVAPR-----DOEFTDK 442
Db 600 EDQEGEYQKLDPEYKAMQDKIAMRGAESISEDMKKQEGTAELSENAKKEVEDTARES 659
Qy 443 AYNLTEAHKALFYNK-----GRNSDFQALDKLERLNDSTKKEKLYVDLLAFIAP 495
Db 600 AEGVVEKSKTPESKRYKCTSGRPEDLOINERPEVLKEDVRYPDDVAPPELATIEN 719
Qy 496 THPERLGPNSQIEYTEDEVRIADLADKYTTSQGYIFDEHD-----IISDEGDAVYTPHM 550
Db 720 SEEDPEKQSORVOISEQAEITQKMDGVGSTTS---FKEEKKPKRFEITQGDKITGKDT 776
Qy 551 GHSNHWIGDSIDSEKVAQAAYTEKEGILPRS-----PDADYKANPTODSAAAIYNR 602
Db 777 NHEGEATEAASENSK-ASDVGTAEKYTEPSSSEYKKDTEBDAEAE-----NSEKTEFIK 830
Qy 603 VKGEKRIPLVAPVVEHTEVKNKGNLIT-----PHKDHYNIKFAMFDHDTY 650
Db 831 VKAE--LENDLAPKAEVYTALENKEDNEVEDTEBDAEVENSEKTEFIKRV---AELGNL 885
Qy 651 KAPN--GYTLE-----DLFATIKYVE---HPDERPHSNDGWNASHEVLGKDK-- 694
Db 886 DAPKAEVYTALENKEDNEVEDTEBDAEVENSEKTEFIKRV---GCTEAEVSKKNAE 942
Qy 695 --HSDPPKKNKKADEPEVEETPAPEVYQVE-----TEKVE-AQIKEAVLLAAYTSSSL 746
Db 943 AVTKEDENMENSIAEALKDVTGQDEIDIDINISDEFORTVELPELEKODIDKNGEDKEL 1002
Qy 747 KANATETLAGL-----RNLFLQINDNNSIMAEKLLALLKGSNPSVSKKIN 796
Db 1003 EVELTEKETSIPDLVEENIT---EKKNEIKOEBE-VSQDENETESISKAPN 1053

```

## RESULT 15

```

D84900
hypothetical protein At2g46240 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84900
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanacker, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402:761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84900
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1043 <STO>
A:Cross-references: GB:AE002093; MID:93702325; PIDN:AA062882.1; GSPDB:GN00139
C:Genetics:

```

A:Gene: At2g46240  
A:Map position: 2

Query Match 4.0%; Score 168.5; DB 2; Length 1043;  
Best Local Similarity 18.5%; Pred. No. 0.22;  
Matches 169; Conservative 108; Mismatches 327; Indels 311; Gaps 38;

```

Qy 40 KREGINABQIYKTIIDQGYVTSHGDIHYNGKY---PYDAMIISBELMKDPNRYLKDED 96
Db 271 KRDDVEASE-----SNEDRRKMONGKIYEFYFIDISMIKSL----- 306.
Qy 97 IVNEKGGYVIAKVDGKYVYVYLKDAHADNVRKBEI-----NRQKQESHQ 142
Db 307 ---IQG-----QDYVEAQONQKNKEEPGGVPIPIFIPSYGRKKDVEASES 348
Qy 143 REGGTP-----RNGCAVALASQGRYTTDDGYITNADIIIDTGDAYIVPHGD 190
Db 349 KESNENHLESCPSDLHRNEGQITQAGK-----EGNEECNVLSDAEKKSVINIPVAN 403
Qy 191 HY-----HYIPKNLSASELAAEAFLSGR----- 215
Db 404 HLOEPRNIPVLSLENHLKPTPEPTKRIAKNEPVSKTEQSSSSSEASKLPPVCLRDPL 463
Qy 216 -----GNLSNSRTYRONSQNT-----SRTNVPYSVNPQTTNTNTS---NN 254
Db 464 PKERNGSKSVSHPRKMEKSKETKIAAPLSKKAESRT--VPEACNVCEADANEMKMAE 521
Qy 255 SNTNSQASQSDIDSLKQLYKPLSORHVESDGLVFPDPAQITSTRTAGVAVPHGDHYH 314
Db 522 GSLNALRTEKGSVES-----NSNLOEESNGEIIIRPCEAKENREOP----- 561.
Qy 315 IYPSOMSELEBRIARIIPLRYS---NHVPDSRPQSPQPPPEPSG----- 360
Db 562 ---AKSTTEEBARIIISMTRGIDVRKWEPIKIKEIA---TYREQGDYKKRIEALFA 615
Qy 361 --POPAPNLKIDSNSSLVSQLYARKVGEGVFEKGISRYVAKDLPESTVYNLESKLS 416
Db 616 STDQHIEKEIYVNGELVAMNLLKLD---AVEGLHPSIREFRKALATE-LSSIQDKLSD 671
Qy 417 KQESVSHITLAKKENVAPRDEPFYDKAVNL---LTEHKAFLFKKGNPSQALDKLLE 472
Db 672 KNSCASAKEKVAQVEIKSQP-SDSPVNLHSQLTTEKNKV----- 712
Qy 473 RLNDSTKKEKLYVDLLAFIPIHPERLGPNSQIEYTEDEVRIADLADKYTTSQGYIF 532
Db 713 ---SDTNLEKVL---LSPEEH-----PMSVLANRTDEK---QAESAAETEBS-- 751
Qy 533 DEHDIIISDEGDAYVTPHMGSHWIGKDSLSEKVAQAAYT-----KEKGIIPSPDA 585
Db 752 -----GUFETLATDSKQATENAAAASSTTIEKIGEVETVYVPGN- 791
Qy 586 DYKANPTGDSAAAIYNRVKGKRI-----PLYRLPVMEHT-----VEV--- 624
Db 792 ---PSADGNGMTVTVTEENKAMAVESLEPIINELPQMEVETENSIIRDENASEVSEA 846
Qy 625 -----KNGLLIIPHKD---HYNNIKFAMFDHDT---YKAPNGYTLLEDLFA- 663
Db 847 ETNSSENNENKRGEDDIYHSEKNVELSELPGVYIDEETQPLSQDPSSTYREGMNTAMP 906
Qy 664 -TIKYVEHPDERPHSNDGWNASHEVLGKKDHSDEPNKKNKADEPEVE-----ETPAE 716
Db 907 KTAQOEETEVDSHNSNKGIGQQTSEPODEKEGS--PETEIVYEQGLETFEVLINBQAP 964
Qy 717 PEVQVETEKVYEAQLKAEVYLLAAYTSSSLKANATETLAGIRNLLQINDNNSIMAE 776
Db 965 PETTEPGISKETKIMEENQFKETMETIVYAGREG--LEVISKLTGRVXSLERKLSHK 1022
Qy 777 KLIALLGKSNPSVS 791
Db 1023 KTOIRRRASKPMVS 1037

```

Search completed: November 13, 2002, 04:11:22

Job time : 55 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 14:49:09 ; Search time 25 Seconds

(without alignments)  
1320.606 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SEIGLYQARTVKEKNRVS.....KLALLKGNPSVSKEKIN 796

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	4.1	1233	YF16_YEAST	P43597 saccharomyc
2	161	3.9	1420	APX_XENLA	Q01613 xenopus lae
3	158	3.8	1118	UBP8_HUMAN	P40818 homo sapien
4	157	3.8	910	DNJM_MYCPN	P73354 mycoplasma
5	157	3.8	1658	YMF7_YEAST	Q03661 saccharomyc
6	157	3.8	2195	SC16_YEAST	P48415 saccharomyc
7	156.5	3.8	1185	MAPX_DROME	P23226 drosophila
8	154.5	3.7	3924	ANK2_HUMAN	Q01484 homo sapien
9	154	3.7	1453	Y373_BOVIN	Q91u23 bos taurus
10	153.5	3.7	1539	Y373_HUMAN	O15078 homo sapien
11	150.5	3.6	1176	YOH8_YEAST	Q08236 saccharomyc
12	149.5	3.6	1175	PTNL_RAT	O63728 rattus norv
13	149	3.6	2869	RBP1_PLAYB	Q00798 plasmodium
14	145.5	3.5	1085	CUT7_SCHPO	P24339 schizosacch
15	145.5	3.5	1240	YNU1_YEAST	P53935 saccharomyc
16	143.5	3.4	900	SYA_MYCGE	P47534 mycoplasma
17	142.5	3.4	875	Z1P1_YEAST	P31111 saccharomyc
18	142.5	3.4	1066	STMS_SCHPO	O74454 schizosacch
19	142	3.4	850	D7_DICDI	P54682 dictyostell
20	142	3.4	1744	TANA_XENLA	Q01550 xenopus lae
21	141.5	3.4	490	MOT3_YEAST	P54785 saccharomyc
22	141.5	3.4	1928	MYSL1_YEAST	P08964 saccharomyc
23	141.5	3.4	2314	PRP2_HUMAN	P23471 homo sapien
24	141.5	3.4	2748	NUM1_YEAST	Q00402 saccharomyc
25	141	3.4	914	IFR42_YEAST	P33936 saccharomyc
26	141	3.4	1332	SPF7_YEAST	P33177 saccharomyc
27	140.5	3.4	516	P54_ENTFC	P13692 enterococcu
28	139.5	3.3	1609	LMG1_HUMAN	P11047 homo sapien
29	139	3.3	1957	YDB6_SCHPO	Q10411 schizosacch
30	138.5	3.3	1703	SNF2_YEAST	P22082 saccharomyc
31	138	3.3	1639	MSF1_PLAEM	P04933 plasmodium
32	137.5	3.3	1790	USO1_YEAST	P23386 saccharomyc
33	137.5	3.3	2175	HMCU_DROME	P10180 drosophila

34	137.5	3.3	2459	1	MAPB_RAT	P15205 rattus norv
35	137	3.3	1176	1	PTNL_MOUSE	Q62136 mus musculu
36	136.5	3.3	679	1	YIS3_YEAST	P40563 saccharomyc
37	136.5	3.3	1183	1	CNA_STRAU	O53654 staphylococ
38	136	3.3	1630	1	MSF1_PLAEM	P04932 plasmodium
39	136	3.3	1636	1	BUD3_YEAST	P25558 saccharomyc
40	135.5	3.3	1053	1	SLPM_BACBR	P06546 bacillus br
41	135.5	3.3	1976	1	MYHA_BOVIN	Q27991 bos taurus
42	135.5	3.3	3358	1	PGCV_MOUSE	O62059 mus musculu
43	135	3.2	1532	1	IGA_NEIGO	P04790 neisseria g
44	134.5	3.2	960	1	YMX6_YEAST	Q04279 saccharomyc
45	134.5	3.2	1124	1	TCF8_HUMAN	P37275 homo sapien

## ALIGNMENTS

```
RESULT 1
ID YF16_YEAST
AC P43597:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Hypothetical 137.7 kDa protein in UGSI-FAB1 intergenic region.
GN YFR016C.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=93400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -1 SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: D50617; BAA09255.1; -
DR SGD: S0001912; YFR016C.
DR KJ Hypothetical protein.
DR KW SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;
SQ
Query Match 4.1%; Score 171; DB 1; Length 1233;
Best local similarity 19.8%; Pred No. 0.1; Mismatches 402; Indels 222; Gaps 42;
Matches 189; Conservative 142;
QY 12 VKENRVSYIDKQATOKTENTLPDESVKREGINMEQVI-----KITD-QGYVTS 61
DB 151 IKETSYNNVAEG-----TENVP--IKESTGIEVGNSTPRKKKKKKKTNNRRGRNS 202
QY 62 HSDHHYHYNGKVPYDAIT--SELLMKDPNYLKDEDI-----VNEVGGVYIKVD 110
DB 203 NPADTTDLSKOSTLSDISLIGIEIYLOEDGS---KNEDIKVINIVODEPVVWEKMDIRTRNE 259
QY 111 GYYVYLYKKAADANV-----RTKEINRKOEHSHREGG--TPRNDCAVALARSQ--- 160
DB 260 SSDKTFDDIVPKNDVDETSKSENNINNEKAHHTLPRENETLNVNBSNAASFHOLEP 319
QY 161 -GRYTTDDGYIFNPASDIETDGDAYIVPGDHYYHYIPKNELSSASELAAEAFISGRGNLS 219
```

```

Db 320 HGLEAGDE----NGAOSTKDVESESLTKMGFNKFNESKHLKAGEKOOTE---SDRDGIS 372
QY 220 NSRHYRONSDMTSPINWVPSVSNP-----GTTNTNTSNSTNSQASQNDI----- 267
Db 373 PSYLANQOKETEELGKEDHVEEQDKDEKCRKELSYNHENHNSHNRNAAGSDSITIPETE 432
QY 268 ----DSLKQLYKPLPSORHV-----ESDGLVDPDPAQ---TISR 299
Db 433 RETYDEDTMGPTKRISDNENKLOHGTNDISVEYEKEEEEEEEENSTFSKVKENVTGE 492
QY 300 -----TARVAVPHGDHHTFYQSOMSELEERIALIPIRYASN----- 358
Db 493 QEAVERNNEVSGTEESTSKGEELMGD-----EKQSEAGER-SSTIEGSAANSAKIS 544
QY 339 --HWVDSRPEQSPQPTPEPSPGPAPRLKIDSNSLSQLYVRKYGEGVEEKGISR 396
Db 545 KDLVLEDEAEAPTOEKKPEYVGEIDIPAPRDD-----VELVEAVENKIIIPDELEVAK 599
QY 397 -----YVFAKDLFSEYVKNLESLSKQESVSHTLAKKENAVAR-----DOEFTDK 442
Db 600 EDEGEQVKLDEPVMKMKDKIMRGAESISEDMKKRQOETAALSNEKAKEVEDETARES 659
QY 443 AYVLLTEAHKALPYXNK-----GRNSDFQALDKLLERLNDESTNKEKLVDDLAFIAP1 495
Db 660 AEGVEVEKSKTPSPKPVYKRCITSGRPEDLIQINRDEPVLAKEDVRVDEYKPEIATTIEN 719
QY 496 THERLGRPNQIETYEDEVRIRIQLADKYTTSQGYIFDEHD-----IISDEGDAYVTPHM 550
Db 720 SEEDPCKSQVQISTQEALETQKMDGVDGSTS---FKEEKKRPREITQEGSKITGKDT 776
QY 551 GSHHWIGKDSLSKEVYAAQATYKKEGILPPS-----PDADVKANPFGSSAAAIYNR 602
Db 777 NHHGATEAASSENSK-ASDVGTAEKYIEPSSSESVKKTDEDAEVE---NSEKTEFIK 830
QY 603 VKGEKRIPLVRLYMEHTYEVKNGNLI-----PHKDYHNKFKAMPDHY 650
Db 831 VKAE--LENDAPKAEVLTAEMLKENEDVEVTEDEAEVENSSEKTEIKVK---AELGNL 885
QY 651 KAPN--GYTLE-----DLFATIRIYVE---HDERPHSNDGNGNASEHVLGRKD-- 694
Db 386 DAKKAEVLTAEMLKENEDVAVATSKEDITKCSAPAETIED---GTCEAEAVSKKDAE 942
QY 695 --HSEDPNKKFADEPVEETPAEPYPOVE-----TEKYE-AOLKAEVLLAKVTDSL 746
Db 943 AVTKEDENMNSKIAEALKVGTGDOEIDIDINISDEFORTVELPELKEKODIKDKGGEKEL 1002
QY 747 KAAATEPLAGL-----RNNLTQIMDNNSIMAEKILLKSGNSSSVSKENIN 796
Db 1003 EYEETEKETSLPDLVVEENIT---EENKIKOEER-VSOLDENETESISKEAPN 1053

```

RESULT 2  
APX\_XENLA  
ID APX\_XENLA STANDARD: PRT: 1420 AA.

AC 001613;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Apical protein (APX).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Ovary;  
MEDLINE=93107151; PubMed=1334959;  
RA Staub O., Verrey F., Kleyman T.R., Benos D.J., Rossier B.C.,  
RA Kraehenbuhl J.-P.;  
RT Primary structure of an apical protein from Xenopus laevis that  
RT participates in amiloride-sensitive sodium channel activity.";

```

RL J. Cell Biol. 119:1497-1506(1992).
CC -I- FUNCTION: IS PART OF A MULTIMERIC COMPLEX WHICH IS INVOLVED IN
CC AMILORIDE-SENSITIVE SODIUM CHANNEL ACTIVITY.
CC -I- SUBCELLULAR LOCATION: Membrane-associated.
CC -I- TISSUE SPECIFICITY: KIDNEY, PROXIMAL INTESTINE, OOCYTES, AND
CC TO A LESSER EXTENT IN THE DISTAL INTESTINE, STOMACH AND EYE.
CC -I- SIMILARITY: SOME, TO HUMAN APXL.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z14997; CAA78718.1; -.
DR PIR: S25517; S25517.
KW Membrane; Sodium transport; Transport.
FT DOMAIN 111 114 POLY-SER.
FT DOMAIN 559 564 POLY-SER.
FT DOMAIN 1048 1051 POLY-THR.
SQ SEQUENCE 1420 AA; 159467 MM; 0DD8B5C11413FBFC CRC64;

Query Match 3.9%; Score 161; DB 1; Length 1420;
Best Local Similarity 19.8%; Pred. No. 0.42;
Matches 180; Conservative 111; Mismatches 318; Indels 302; Gaps 44;

QY 42 EGINAQIYIKITDQGYVSHGDY-----HYNGKVVY-----DAISFE 82
Db 26 ERISPVRSMTTLVDASVYSSFSGSSTYPEYQNFQHDGCHYNDQLSYMDSEYVALIYNS 85
QY 83 LLMKDPNYKLKDEDIVNEKGVYIKVDK-----YYUULLDAHADNVKTR-- 129
Db 86 LLDKDGYY---NDIYSE-HGSSKVALSGRSSSLCSMDTTSVHRKSPAKLDNYVNLDS 140
QY 130 -----EELN-RQKQESHQREGGTPRND--GAVALARSQGR-----YTTDDGYIFNA 173
Db 141 EKNIVGDPINMKHKQRPNHKAVGLQNRSPGTGINSIQEKENQLYNPSNMEIKDNFYGRS 200
QY 174 SDLIETGDAIYVPHGDHN-----YIPKNELASLSLAALAEFLSGNGLSNSR---TYR 225
Db 201 LDVLQADGQ---IMTQDSYTONALYFPQNPQOYR---NTQYPGANRSKQOFKYNVQ 253
QY 226 RONSNDTSPRTNWPVSNSNGTNTNTNSNSNTNSQASQNSNDISLKLQLYKLPLSORHVE 285
Db 254 KSNHEENTERDG--PYLTRGQF-----VQGYASDVATSKNI-----RSLK 294
QY 286 SDGLVDPQAQITSRTARG---VAVPHGDHYHF-----IPYSOMSELEERIALI-- 331
Db 295 KSA-----SGKIYVHDSQSGCWMKPGKOTPFSEGTITDMQVNDREQMDIRKSLSTRA 350
QY 332 --PLKRYSNHWVDSRPEQSPQPTPEPSPGPAPRLKIDSNSLSVQLYVRKYGEGYVF 389
Db 351 SOSLYYESMEDV-----SGPPLKAAAMSKNVDLTLS----- 381
QY 390 EEKGISRYVFAKDLFSEYVKNLESLSKQESVSHTLAKKENAVAPRDOEYDKA-----Y 444
Db 382 -----FQKDAIVKSTIPLLSQLOQEKCKSHPL-----SDLNCKITKASTPMY 425
QY 445 NLITTEAHKALF-----XNGRNSDFQALDKLLERLNDESTNKEKLVDDLAFIAPTRPE 499
Db 426 HLAGSHSAFIAPVHNTNPAQOEKCLKLESKTLEBMNNS-----VLQISEPRDNH 476
QY 500 RUGKPNQIETYEDEVRIRIQLADKYTTSQGYIFDEHDIISDEGDAYVTPHMGSHWIGKD 559
Db 477 KLPKKNKS-----LTQLADLHDSVEG-----GNSGNLNSABE 508
QY 560 SLS-----DKEXYAAQATYKKEGI---LPPSPDADVKANP----- 591
Db 509 SLKNDYIEKLVAQKQKVLRKTSFKRKDLQMSLPCRFKLPKRPRTIDHFRSYSSSSANEE 568

```



```

RESULT 4
ID DNM_MYCPN STANDARD; PRT; 910 AA.
AC P75354;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DI DnaU-like protein MG200 homolog.
GN MPN19 OR MP035.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.,
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AE000004; AAB95683.1; -.
DR HSP; P25685; 1HDJ.
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
DR Hypothetical protein; Chapterone: Complete proteome.
FT DOMAIN 4 73
FT SEQUENCE 910 AA; 100190 MM; 125DDE37D2D221A7 CRC64;
Query Match 3.8%; Score 157; DB 1; Length 910;
Best Local Similarity 19.0%; Pred. No. 0.38;
Matches 161; Conservative 127; Mismatches 305; Indels 254; Gaps 44;
QY 66 YHHYNGKVVYDAIISEEL-----LMKDP---NYKLKDEIVNEVGGYVIKVD--GKYY 114
DB 34 YHDKRKAPDAAOITFAIEIENDVLSNPKRKANYDKYGHGVDN-EGGFAFOADVDFDSEF 92
QY 115 VYLKDAAHADNVATKEEINROKQESHQREGGTPRNDGAVALARSGRTTDDGYTFNNA 174
DB 93 EEIEKSGARDNL---SESNTKKKEKTKKKGW-----FMGSKQDEISTDTTEY--A 140
QY 175 DIIEDGDGYIYPHGHYIYIKNELSASELAAAEFLSGRLNSKRYRRONSNTSR 234
DB 141 DV--DAGLEDYPOSIDYPPDIPDVARIIEVQS--AYADIPDVAGMDEQNAEYANSA 197
QY 235 TMMVPSVNPQTNTNTNSNTNSQASOSNDISL----- 271
DB 198 SETIPV-----DAGLAEFNTSSAAPQASDWEANTGMPGYCPDAGCEMNMKGFTDRA 251
QY 272 -KQYKRLPLSRHVESDGLVFPDPAQTSRTARGVAVPHGDHYHFIYPSOMSELEERARI 330
DB 252 GQVMVLEETPESSVSNDETTTSDAVTAAT-----VEETDQ----- 288
QY 331 IFLRKSNIWVDSRE-----QPSQPTP-----EPSPGP-----QAPNLKI 369
DB 289 -----DSWTNSAPEPVDETPVELQPEPEPIITLISSEVPEAPASVIEPTPEIRE 341
QY 370 DSNSSLVQLVKKVGGVYEEKGISRYFAKDLPESETYKNLESKSKSRESHITLAK 429
DB 342 TTSAVEMDASVAK---ADVSDADATNEPTEDTISEPQ--ETDAALAEIINH-TADL 394

```

```

QY 430 ENVAPROEFDYKAVNLTLE--AHKALFYXKGRNSDFOALDKLERLNDSTNKEKLYVD 487
DB 395 E--PAEV-----SATNDLDQDVYEKVFSEPESTVDTATDPVDAQETTSNGFF--N 446
QY 488 LIAFL-----APITPERLGRKNSOI-----EYDEKVIKQALA----- 521
DB 447 FSSFVLSDGNPNPQPTTHHEEADAAPETVDETSGETSPAEVYIAESTVLETAIEINNP 506
QY 522 -----DKYTSIDGYIFDEHDIISDEGDAVVPFHMGSHWIGKDSLS 562
DB 507 AIFVEEYLDPTKTYVDKLDPELVAKPTVSDSENSVAPPEVPAEGQTF--W--KPAIS 563
QY 563 DKEV-----AAQAYTKKGLPPSPDA-----DYKANPTGDSAAA-----IY 600
DB 564 EEIEIPLTAVEPASETQTLIAEDVTSPTVPTAFAPISINAVPTAPVETFEAAVFLK 623
QY 601 NRKGEKRLPLV-RLEPMVEHVEYKNGNLIIPHKDHYNIKFAMFDHTYAPNGTYLE 659
DB 624 EAKIEAQLPLVPTVEQIDGT---DPSIL-----TQWDEYLEKTRK--LPH 665
QY 660 DLFATIKYVVEHDERPHSDGKGNASEHVLGKHSE--DPNKNKKADEEVEETPAEP 717
DB 666 KLELT-----EQLP-----FIVKTDQFEIYDPA-----LDEHNVNLIYTE- 700
QY 718 EYPOVETKEVQAQKEAEVILAKVTDSSLIKANATETLAQLNNLTLOIMDNNSIMAEAK 777
DB 701 HVPQICF--LNNQDKIR-YTRKLVDPQVYTTT-----SITLEVLGSHKSQTEA-- 748
QY 778 LIALKLG 784
DB 749 -IAIFKG 754
RESULT 5
ID YM67_YEAST STANDARD; PRT; 1658 AA.
AC 003661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-1997 (Rel. 35, Last annotation update)
DI Hypothetical 187.1 kDa protein in G0A1-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=S288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: Z49809; CAAB9934.1; -.
DR EMBL: Z49939; CAAG0190.1; -.
DR SGD: S0004832; YMR219W.
KW Hypothetical protein.
SEQUENCE 1658 AA; 187137 MM; 3893F968305A757D CRC64;

```



```

Db 29 HSEBPLEPESTINSS-----FNDDSVNRTESDIASKSD-VPPVS-----S 68
OY 248 NNTS-----NNSNTNSQASOSNDI-----DSLKOLYKLP 279
Db 69 SNISIPANETOLEIPDTOLHKLKLNDSQDHDITADSNDLPNSIYEHDSVTIOT-KPAM 127
OY 280 SORHVESDGLVDPQAITSR-----TARGVAAPHGHDHFIFYSOMSE----- 322
Db 128 SOGEYEBT-----AAHLSRNPISLDVAGELHNNHNDHTOKIAVSAVEEDSFNEEGENDH 181
OY 323 --LEERIAIITLTKRSNNWVDSRPEOSPPPTPEPSGPGAPAPRLKIDSNSLSVQLV 380
Db 182 STIISLNDATPSQY--NHFLPSD-----GNLSPELSSGDPTHNVPGLTGDN----- 228
OY 381 RAYGEGYEEKGRYFAPKD-LPSETYKNLESKLSKQESYHTLTAKENAVPADOEF 439
Db 229 -EINDEYONDEIS--LNANVLDELDELSEDERKLETHYS--TEEKKODIA--DOET 281
OY 440 YKAVNLTLEAHKALEFXNGRNSD--FOALDKLERLNDSTNKEKLVLDLAFAPITH 497
Db 282 AENLFTSTPESENKIRNSGDDTSMLFQ-----DDESQKVPWEDVKKDF-----H 328
OY 498 PERLKPNSQIETDEVR-----IAQLADKYTSDGIIPDE--HDIISDEGAYVTP 548
Db 329 NENTNTOESAPNTDRODGYEGNEALKKSESCTADERSYSEETSEDFHGHDKQVVE- 387
OY 549 HNGSHWICKDSLDEKRYAQAAYTKERKGLTPSPADAV--KANPTGDSAAAIYNNKGE 606
Db 388 --GONDFCTKNLENESQKLMGEGNHK-----LPLSADIIIEGCKDQDADLEDTFOSSD 441
OY 607 KRIPVLRLPYWVEHTVEYNGNLIIPKDHYNHNI-----KFWN---FDDHYTKADN 654
Db 442 -----LGEVLPWEMSTDNADVTSKSQKHEDLFASGNDKLEPVEVSDGEVSSGKTEN 494
OY 655 G-VTLEDLFATIKY--VVEHDERPHSNDGMNASEHVGLKHSDEPNKFNKADEPVEE 712
Db 495 SNOTSTERTAECKRFSFLENDDDLDDDSFLASSE---EEDTVPTDNTJTLTSKPVVE 550
OY 713 TPAPPEVPOVETKEVAQLEAEVLLAKVT 742
Db 551 KKASRRKPIIEE---EAGMRQOVHETNTT 577

RESULT 7
MAPX_DROME STANDARD: PRT: 1185 AA.
AC P23326: 09V9S1:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 205 kDa microtubule-associated protein.
GN MAP205 OR CG1483.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman Y.-H.C., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broststein P., Brottlter P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies L.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclele J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector A.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF
CC MICROTUBULE ASSEMBLY AND INTERACTION.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH CYTOPLASMIC MICROTUBULES AND
CC WITH THE MITOTIC SPINDLE.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; C2, B3 AND J5 (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- MISCELLANEOUS: PHOSPHORYLATION OF VARIOUS SERINE RESIDUES MAY PLAY
CC A REGULATORY ROLE. THE BASIC DOMAIN CONTAINS NUMEROUS SEQUENCES
CC THAT MATCH KNOWN CONSENSUS SEQUENCES OF SEVERAL DIFFERENT PROTEIN
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54061: CAA37996.1: -
DR EMBL: AE003780: AAF57214.1: -
DR PIR: A36685: A36685.
DR Flybase: FBgn002645: Map205.
KW Microtubules; Alternative splicing; Phosphorylation.
FT DOMAIN 1 784
FT DOMAIN 785 1124
FT BINDING 1125 1185 ASP/GLU-RICH (BASIC).
FT BINDING 745 977 ARG/LYS-RICH (BASIC).
FT VARSPLIC 557 578 TO MICROTUBULES (POTENTIAL).
FT VARSPLIC 650 703 MISSING (IN ISOFORM B3 AND ISOFORM C2).
FT VARSPLIC 704 704 MISSING (IN ISOFORM C2).
FT SEQUENCE 1185 AA: 126669 MM: 4784222ECC03F70 CRC64:
SQ
Query Match 3.8%; Score 156.5; DB 1: Length 1185;
Best Local Similarity 20.1%; Pred. No. 0.58;
Matches 163; Conservative 111; Mismatches 304; Indels 231; Gaps 36;
OY 122 HADNVTKEEI-NRQKQ-----EHSQHREGGPRNDGAVVALASGGRITDGGYIF 171
Db 4 HEDNQLDNYLQNRLESIQICGAGEHNPILADATGGNGCAPGAPSK----- 52

```



Qy	172	NASDIIEPTGATYVPHGDHHTIPKMLASSELAAAE-----AFUSGRGINS	221
Db	53	---SDVEDEED-----EEKKYI--HEVROSEKLOOEKPLTKEGTNGFGPGR-----	95
Qy	222	RTYRNRNSDNTSRITNW-----VPSVNSGCTNTNTNSNSNTNS-----QASQND	266
Db	96	---DSDNVHGNGAAAVNLYKEEDVEYLYKNGDSTNSNTTSTTSTDEYVARQAEPNQ	149
Qy	267	IDSLLKQLYKLPQSQRHVESDGLVFEDPAQITSRTPAGVAVPHG-----DHYHF	314
Db	150	LPEQLQO-----QQQLSESQGVHEDDPEQDEDEHSSVATYGTSHSLENNSSPLDQEEV	202
Qy	315	IPYISQ-----MSELEERARIILPLRYSNHNVDPDSRQEQSPQPTPE	356
Db	203	VMAVATYGOEOLVDPDNKENSIFYKNNLEENHSQNP--NAVAVPPVGGSSSSPLPAE	259
Qy	357	-PSPGPAPAPMLKIDSNSSLVSQLVKRYGEGEYFEEKGISRYVFAKDLPSBYVNLSEKL	415
Db	260	DPLPEVQGRPLPGGTLDDLVAESPRK--EPARINMDGIA-----VPPERREDIADM	310
Qy	416	SKQESVSTTLTAKKENAVAPRQDEYDYKAVNLLTEAHKALFYXNGRNSDFQALDKLERLN	475
Db	311	R-----PHELEQESDPTGAGHLEMQ-----LLNGIGTADOALRDVLDPGR	351
Qy	476	DESTNKEVLVDLLAFIAPITHPERLCKPNSQIE-----YTEDEVRIQLADKY	524
Db	332	ETSDVME--LPIDQVPPNDADIKOSITVAEHNSIIDINSVQPLPIQTCDKELHYEKE	410
Qy	525	TTSDGYITDEHDIISDEGDAAVYTPHMGSHHIGKDSLDEKVA-----AQAYTKE	575
Db	411	HVKSASPTREELQFQSD-----PNNQESHTEPLNTEQDDPMQASFYLEHNSQRAQBEQOE	466
Qy	576	KGILPPSPDADYKANPT--GDSAAATYNRXKGEKRIPVRL-----	614
Db	465	QMOU-PAECSDIFADOSLILDTSTAP--QLSEADSEVAKLELSQAGIYDITPSSLSS	520
Qy	615	--PYMVEHTVEVNGNLIIPKHKYHNIRKAWFDHTYKAP---NGYLLDEDFAPIKY	667
Db	521	TAEKHLVDTELVETETLDP--ESHFPGVY-----SSQAPLOLFGSKHTLPLSTHSCKH	572
Qy	668	YV--EHPDERHSDNGMGNSEHVLGAKKDHSEDPNKNFKADEPEVETPAPEPEVQVETE	725
Db	573	RVASPQND---ENAVFESVSGYETQNFDEISSPPEGINFAQPF--TPAHIVIEQANTM	627
Qy	726	KVEA-----QKEKEVLLAKYTTDSS---LKANATETLAGLRNNITLQIMD-----	767
Db	628	MEDVGMDIPASEDFACIDKVASSSNEVEDHRSQOAFVKEELLHPGVGVAVQAVENLGT	687
Qy	768	NNSIAAEKRLALLKGSNP--SSYSKKRI	795
Db	688	EKNFVEEERLPIYSDEIPLSSASKERL	716
RESULT 8			
ANK2_HUMAN			
ID	ANK2_HUMAN	STANDARD;	PRT; 3924 AA.
AC	Q01484; Q01485; STANDARD;		
DT	01-APR-1993 (Rel. 25, Created)		
DT	01-JUN-1996 (Rel. 34, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).		
GN	ANK2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI.TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).		
RC	TISSUE=Brain stem;		
EX	MEDLINE=91302466; PubMed=1830053;		
RT	Otto B., Kunimoto M., McLaughlin T., Bennett V.,		
RT	"Isolation and characterization of cDNAs encoding human brain		









```

FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BC6F CRC64;

Query Match 3.6%; Score 149; DB 1; Length 2869;
Best Local Similarity 17.2%; Pred. No. 4.9;
Matches 166; Conservative 148; Mismatches 336; Indels 314; Gaps 42;

QY 24 KATQKTEMLTDEVSKEGCIATQ-----IVIKITD-----QGT---VTSH 62
DB 1702 KESYEKNLETFVQNEISR---INVEBSLTDIDKKTIDIDNLLKMKQYEEGLQIKEN 1758
QY 63 GCHYHYNGKVPYDAIISPELLMKDN-----YKLKDIYNEVGGVIVKDGKYYV 116
DB 1759 AD-----KKSXNPELVGSILNALDPSTIFIKLIKETDMGDLK-NGVAMNEIHGEF 1812
QY 117 LKD---AAHADN-----VTRKEBINRQKQSHQRE----- 144
DB 1813 TKSYNLIEFHLNADTVSVPFEKAQSLRELAKKEEHLRREBEAFLINDIKKVESLKL 1872
QY 145 -----GTPPRDGAVALARSGRTTDD-----GYTFNASDIIDETG 181
DB 1873 LKEMKKVSAEYEGMKRDSYVSQIVQDKTIVDELKTLNDISEGSSVYNVVSIVKVK 1932
QY 182 DAYIVPHGHHYIPKNEHSASLAAEAFLSGRNLNSRTYRONSNTS---RTNV 238
DB 1933 ES---KHADYRR--DANSYTESKVTIANYFLSDEAKISSGMEAMKSNKFTDLELTF 1987
QY 239 PSVSNPGTTNTNTSNNSN-----TNSQASQSDIDSLKQYKPLSRHVESDGLV 290
DB 1988 SVTASNELLKTIQDSQNDVQIKERESEQLADATDIYVVK--LKNENKMLEKKNE 2045
QY 291 FDPAQITSTAGVAVPHGDHYHFIYSOM---SELEETIARIIPRYSNMHWPDSPRE 347
DB 2046 EYVSEKVPREALRLSQVESIRCFHEFHRLDNTTELEMLKMKVTL-YR-----DKKSE 2098
QY 348 OPSPOPTPEPSGPQAPAPMLKIDSNLSIVSQIVRKVGEGYVEEKGISRYFAKLDPSST 407
DB 2099 RES-----GLQEMEN-EMNTYSNSTQL-----BEIVVSAGSKSDIDIKLEKSNNE 2143
QY 408 VNLESKLSKQBSVSHTLTAKKENAVPRDQFYDKAVYNLLTEAHKALFYNNKGRNSDQAL 467
DB 2144 MNISIKIS-----TIDSKYIEMNSTIDELYLKLGKN--COAHMISLISYTN--MKT 2191
QY 468 DLLERLNESTNKEKLVDDLLAFLAFLPIHPERLGRKPNQIETDEVRITADLADYTTIS 527
DB 2192 SKKLIMINKENTKCV-----YIKDSSST 2219
QY 528 DGYI-----FDEHDIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQVYTK 575
DB 2220 DGYVEELKFGYSKLFSSASEIYQADITYSVNPAKHE---KESNATARDIKKELYLPH 2275
QY 576 KGLPPSPADYKANPTG-DSAAAIYNRVYKGR-----IPLVRLPYMVBHYEV-- 624
DB 2276 Q-----NSDISIVEGVQNMALYDKLINEKREMDLEYRINSETKIKOM-EHSTDYVK 2327
QY 625 -----KNONLIIP----- 632
DB 2328 PMIELHKGNETNNKSLLEKEKKLSVNDHMSMEAKMKGKLYTPESVQNMNNIYVI 2387
QY 633 -----HKDHY-----HNKFAMFDDHT-----YKAPGYTLEDLFATI 665
DB 2388 EAEVKTLEIDIDYGNVYIVEEHKKQFSIILDRTNALMDDIIEIKKENNYL----- 2440
QY 666 KYVEVHPDERPHS-NDGWNASEHYLKKGDHSDPNKRFKADEEVEETPAEP----- 717
DB 2441 ---MEVNTETIIRVNDYIKRTINKLIVQAKTEYQLEIKQDDMDLQNTFLKAVSTIIEYF 2497
QY 718 EYPOVETEVEVAQLKEAEVLLAKVTD--SLKANATEITLAGLRNNLTLOIMNNSIMAA 775

```

```

DB 2498 ENVKKKKEIINDYEQERLL-KIGBHLDEIKRNVETLTSYEIDOKMEMSKN-LLEKK 2555
QY 776 EKLL 779
DB 2556 SKMM 2559

RESULT 14
CUT7_SCHPO
ID CUT7_SCHPO STANDARD; PRT; 1085 AA.
AC P24339;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Kinesin-like protein cut7.
OS CUT7 OR SPAC25G10.07C.
OC Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91015362; PubMed=2145514;
RA Hagen I., Yanagida M.;
RT "Novel potential mitotic motor protein encoded by the fission yeast cut7+ gene."
RL Nature 347:563-566(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkhardt G., Aert R., Robben J., Grynoprež B.,
RA Welljens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S.,
RA Gabell K., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motlier S.,
RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potshkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880(2002).
CC -!- FUNCTION: COULD BE A SPINDLE POLE BODY MOTOR. ON TRANSITION FROM
CC G2 TO M PHASE OF THE CELL CYCLE. THE SPINDLE POLE BODY DUPLICATES;
CC THE DAUGHTER POLE BODIES SEED MICROTUBULES WHICH INTERDIGITATE TO
CC FORM A SHORT SPINDLE THAT ELONGATES TOSPAN THE NUCLEUS AT
CC METAPHASE. MUTATIONS AT CUT7 BLOCK SPINDLE FORMATION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BINC
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: X57513; CAA40738.1; -
DR EMBL: Z70691; CAA9436.1; -
DR PIR: S14032; S14032.
DR HSP: P17119; 3KAR.
DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin_1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW Mitosis; Cell cycle; Phosphorylation; Repeat.
FT DOMAIN 70 435 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 436 604 COILED COIL (POTENTIAL).
FT DOMAIN 715 740 COILED COIL (POTENTIAL).
FT DOMAIN 897 955 COILED COIL (POTENTIAL).
FT NP_BIND 159 166 ATP (BY SIMILARITY).
FT REPEAT 987 998
FT REPEAT 999 1010
FT MOD_RES 1011 1011
FT CONFLICT 34 61 PHOSPHORYLATION (BY CDC2) (BY
FT SIMILARITY).
FT SASENPKRRREPITIDGYPRSDTNSPT ->
FT LRAILGNDVSLTLTL (IN REF. 1).
SQ SEQUENCE 1085 AA; 122133 MW; 5669277875559D58 CRC64;

Query Match 3.5%; Score 145.5; DB 1; Length 1085;
Best local Similarity 18.6%; Pred. No. 2;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATQRTENLTPEVSKREGINEA-----QIVIKI--TDQGVYSHGDHYHYN 70
DB 159 GGTGTGTYTMSGD-LSDSDGILSEAGLIPRALYOLFSSLDNSNGEAVK-CGYEELYN 216
QY 71 GKAVYDAIISSELLMKRPNTYKLED-----YVNEVKGVIYKDGKYYIYKAAADN 125
DB 217 EEL-RDLVSEEL--RKPARVEEDTSRGVNYIGIEESY-IKNAGQGLRLRGSHRQ 272
QY 126 VRTKEINRQKQHS-----QHREGTFRNDGAVALARSGQRYTDDGYFNASDIHEDT 180
DB 273 VAAKCKNDLSRSISITITLHKRVYSSGMDETNSLTINN-----NSDDLK-- 319
QY 181 GDAYIVPHGHYHYIPNELSASELAEEFLSGRNLMSNRT----- 223
DB 320 -----ASKLHMVDLASENT--GRSGENKRAKETGMINOSLTILGRVI 361
QY 224 -----YR-----RONSMTSRKNWVPSVSNPCTTNTNNSNNTSQAQS 264
DB 362 MALVEKAHHIPRESKLTRELLQDSLGGKTKTSMITVVS--STMTNLEETSTLEYAARA 418
QY 265 NDIDSLKQLYKLPESQHVESDGLVDFDPAOI-----TSRTARGVAVPHGDHYHFIFYSQ 319
DB 419 K--SINKKQNNQVLRKYLINDVLIDIERKLDNATRKKNQVYLAESTYKELMDRVQ 475
QY 320 MSEL-----EERIAIIPLRYRSNHWPDSREEDSPQTPPEPSPQAPAPLKI----- 369
DB 476 NKDLCEQARKLEVLDLNYS-----SREQLQYVSKSNQEHKEVEALQQLVNSSTE 529
QY 370 -----DNSNSLVQVLRKVGEGYFEKGISSRYFADLPSEYKKNESKSKQESVHT 424
DB 530 LESVYSENEKIKNELVLEIERKKRYTNEAKITTVADLSQYRESKEYIASLVEKIDRT 589
QY 425 LTAKEVAVPDDQEFYKAVNLT-----LNDSTNKEKLVDDLAFIAPIT 496
DB 590 ERNNKEN-----ENNFWMLKRLNLTMLRSPFGSTDETNGYFTLLNDNASKEELLNTHSN 645
QY 452 KALFYKNGRNSDFOALDKLER-----LNDSTNKEKLVDDLAFIAPIT 496
DB 646 QLLISMRTKITEHFQSDLEALQASRSCAVPNSSDLIVSELKDSKNSLDLALHESLQDIS 705
QY 497 -HBERLGRPS-----QIETTEDF-----VA--IAQL 520
DB 706 MSSOKLNGISSELLIELQDKKESYROLVQELRSLYNIQHTHESQKELMWGYRNDIDL 765

```

```

QY 521 ADKYYTS--DGYPDEHDIISDEGDAYVTPMGHSHWIGKDSLDEKVAQAAYTKEGI 578
DB 766 VKTQTSIANDDII-LSDYISDQKSPESKQODLIANGKIVSNFLOBQNSLYTK----- 820
QY 579 LPSPDADVKANPTGDSAAAI--YNRYKGEKRIPLVR-----LPYVEHTVEVK 625
DB 821 -----ADILSHLNDPNSINIRKANETIMNNSRSEFLRRAASQAEIVGANKERIOKTYE-- 872
QY 626 NGNLIIFPKDH-YINIFAWFDDHTTKAPNG-----YLEDLFATIKYVEHPDERP 676
DB 873 NGSQLDLSKSKRAHSRSRSMVHDCLALAESQKGVNLEVOITDLRLQVK-----E 923
QY 677 HSNQWGNASAEHVLGKKDHSDDPKNKEKADEPVEETPAPEPEVQVETEKVEAOLEAEV 736
DB 924 HSEDNTEKHQQLDLESLVGNDNL-----IDSIRPHLEQ----- 962
QY 737 LLARVTDSSLK-----ANATETLAGLRN-----NETLOIMNNSIMAEE-----EKLLA 780
DB 963 ---KITDHVLEKGTSLANHTNELLGIDESLCNLETTIEDTSLVKLETTYGDTPSKRELPA 1019
QY 781 LKGSNPSVSKEKIN 796
DB 1020 TPSTWRDSSLIKETTN 1035

```

```

RESULT 15
YNTL_YEAST
ID YNTL_YEAST STANDARD; PRT; 1240 AA.
AC P53935;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 141.5 kDa protein in YPT53-RHO2 intergenic region.
GN YNL091W OR N2231.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=96367601; PubMed=8771715;
RA Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;
RT "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
RT chromosome XIV that includes the ypc53, trnAlau and gsr m2 genes and
RT four new open reading frames."
RL Yeast 12:599-608(1996).
CC -!- SIMILARITY: TO S.POMBE SPAC29E6.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X65811; CAA59826.1; -
DR EMBL: Z71367; CAA95967.1; -
DR SGD: S0005035; YNL091W.
KW Hypothetical protein.
FT DOMAIN 756 761 POLY-GLU.
SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D265822D5778 CRC64;

Query Match 3.5%; Score 145.5; DB 1; Length 1240;
Best local Similarity 19.8%; Pred. No. 2.4;
Matches 148; Conservative 85; Mismatches 236; Indels 277; Gaps 35;

QY 220 NSRTYRQNSDNTSRTMWVSVPSTGTTNTNTSNTNSQAQSNIDISL-----KQLY 275
DB 4 NSKSKRRKN-----KSKQHKKKNGNSDPEQISINPTQLVPRMEPELY 44

```

Qy	276	KLPLSQRHVEDGLVFPDPAQITSTAGVAVPHODHNLPIYSQMSLEIRARII----	331
Db	45	-----HTESD-----YPSRYKAPMGDIVE-----PINTDDKKRTTNLTHNKK	87
Qy	332	-----PLRYR-SNHWWPDSRPEOPSQPRPEPSGQAPAPNLKIDSNSLSVQLVR--	382
Db	88	SMDASSLAFLFLDSHW-----ELSPD-----EKKTLIRKEE--VENVIRNYQ	130
Qy	383	-----VEGVFEFEKGISR-----YFAPKLBPST-----YANLE-SKLSKQ	418
Db	131	DDHSCSCVCGRRHLAMDQEMERYNTLYAMDCKDKDETPNKRHLGIIKELDISKNOQO	190
Qy	419	ESVSH-----TLAKENAVAPROEFDYKAVN--LTREAHKALF	455
Db	191	NDLSTKEEYVKNFLSSSTVGSLSKEEVLHFRKQLSQBOAHNETADNTSLLEENLNTH	250
Qy	456	XNKR--NSDPQALDKLLERLNDSTNKK-----LVDD	487
Db	251	INKTSSEISANFNSVD--BELQOKYENFTKPTISSHPKIAEEYVOKMMYPNIRALTD	308
Qy	488	LL-----AFLAPITHPERIG-----KNSOLEYEDHVRIALAD-----KYTT	526
Db	309	LMSNGGCFVLAIEDVPADGOIQASKDDSI--TEDASSTLDLPKEFTMYLHSGKPLT	366
Qy	527	SDGYIFDEHDIIIDEGDAYV-----PHMGSWIGKDSLSDKEKYAAQAYTK	574
Db	367	EDEVADLORNAIEHMTAYDPAASKFKDVSQLEKELFTFRMSGDKKSPFELLIIQSRNK	426
Qy	575	EKGILPRPDA-----DYKANTGSAAYI-----NNVKGKRI	609
Db	427	FDELGEFVLATLSSCPSOSSKOTSLDTSIYIDDEDEEDYDYSEYAEDESEVSEYEGI	486
Qy	610	PLVRLPVWEHTVEVKNGNLIIIPKRDYHNIKFAMFPDHTYKAPNGYTLDELPAITIKYVY	669
Db	487	EAVEKP---EHD-EKSNIGIRETTLHLSY-----DHDKKQN-----	517
Qy	670	EHPDERPHNSNMGNSAHYLGKDH-----SFDPKNKADEBEPEEETPAPPEYQV--	722
Db	518	HPHHHYHST--STHSDELSEEEYISDIELPHDPKHHFRDDILIDGDEPEEDEN	573
Qy	723	-----ETEVEVAQLAEAVLLAKYVDSLSKANNATETLAGLRNNNTLIOMD	768
Db	574	EGDDEEDTYDSDGDEDTREBLEGRKLIDAITKLLQSSIMASYNEKQA-----DN	622
Qy	769	NSIMAEAKLLALLKGSNPSSVSKK	794
Db	623	NRL-----KLQLEEEKKRKKREEE	643



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 12, 2002, 15:27:09 ; Search time 83 Seconds  
(without alignments)  
1976.066 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKENNRVSY.....KLIALKGSNPSSVSKKIN 796

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4163	100.0	802	16	Q97QM8	Q97qm8 streptococc
2	4163	100.0	816	2	Q9AHT9	Q9ahy9 streptococc
3	3236	77.7	844	2	Q9AG74	Q9ag74 streptococc
4	2797.5	67.2	819	16	Q97QM9	Q97qm9 streptococc
5	2786.5	66.9	819	2	Q9ANY3	Q9any3 streptococc
6	2772	66.6	839	16	Q9ANY2	Q9any2 streptococc
7	1246	29.9	1039	16	Q9ANY1	Q9any1 streptococc
8	949	22.8	825	16	Q99XV4	Q99xv4 streptococc
9	944	22.7	825	2	Q93GT5	Q93gt5 streptococc
10	934	22.4	822	2	Q9ZHG7	Q9zHg7 streptococc
11	679.5	16.3	289	2	Q9AE21	Q9ae21 streptococc
12	223	5.4	792	16	Q99Z76	Q99z76 streptococc
13	197.5	4.7	1078	5	Q963T1	Q963t1 plasmodium
14	192.5	4.6	1390	5	Q770J3	Q770j3 dictyostel1
15	178	4.3	565	5	Q15754	Q15754 dictyostel1
16	176.5	4.2	1271	5	Q25860	Q25860 plasmodium

17	175.5	4.2	1236	5	Q9GTX2	Q9gtx2 plasmodium
18	172.5	4.1	5458	5	Q9U459	Q9u459 plasmodium
19	172	4.1	1795	2	Q9LCJ9	Q9lcj9 staphylococ
20	172	4.1	2478	2	Q9LCH2	Q9lch2 staphylococ
21	172	4.1	2481	16	Q99GR6	Q99gr6 staphylococ
22	170	4.1	2478	2	Q9RL69	Q9rl69 staphylococ
23	168.5	4.0	1043	10	Q82345	Q82345 arabisdopsi
24	166	4.0	1826	5	Q97255	Q97255 arabisdopsi
25	166	4.0	3263	5	Q917U3	Q917u3 arabisdopsi
26	166	4.0	16215	5	Q9NFS3	Q9nfs3 drosophila
27	165	4.0	891	16	Q99TD3	Q99td3 staphylococ
28	164	3.9	891	16	Q99TD3	Q99td3 staphylococ
29	162.5	3.9	2647	5	Q9U4X0	Q9u4x0 plasmodium
30	162	3.9	2703	5	Q9VPL9	Q9vpl9 drosophila
31	162	3.9	2910	10	Q9END5	Q9end5 arabisdopsi
32	162	3.9	5322	5	Q9NT64	Q9nt64 drosophila
33	159.5	3.8	1129	16	Q8XLL1	Q8xll1 cistridium
34	159	3.8	1810	5	Q20456	Q20456 caenorhabdi
35	158	3.8	1650	5	Q77328	Q77328 plasmodium
36	156.5	3.8	1495	10	Q9LZA8	Q9lza8 arabisdopsi
37	156	3.7	519	10	Q9SIC8	Q9sic8 arabisdopsi
38	155.5	3.7	1038	3	Q9P6I9	Q9p6i9 schizosarc
39	155	3.7	869	4	Q9NIF8	Q9nif8 homo sapien
40	155	3.7	920	4	Q14673	Q14673 homo sapien
41	155	3.7	1734	5	Q9UOL7	Q9uol7 plasmodium
42	154.5	3.7	1338	5	Q77306	Q77306 plasmodium
43	154.5	3.7	1463	2	Q86919	Q86919 staphylococ
44	154.5	3.7	1516	5	Q96154	Q96154 plasmodium
45	154	3.7	749	5	Q02602	Q02602 plasmodium

## ALIGNMENTS

## RESULT 1

Q97QM8 PRELIMINARY; PRT; 802 AA.  
ID Q97QM8  
AC Q97QM8;  
DT 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Conserved domain protein.  
GN Sp1175.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
CC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TiGR4.  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Rauden D.,  
RA Holtzapfel E., Khouli H., Wolf A.M., Uitterlind T.R., Hansen C.L.,  
RA McDonald L.A., Felblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RL Science 293:498-506(2001).  
DR EMBL: AF007418; AAK75284.1; -  
DR TIGR; Sp1175; -  
KW Complete proteome.  
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1833 CRC64;

Query Match 100.0%; Score 4163; DB 16; Length 802;  
Best Local Similarity 99.9%; Pred. No. 2e-222;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTNLTPDEVSKREGINAQIVIKIRIDGGVYT 60  
|||||

```

Db 7 SYELGLYQARTYKENVNSYIDGKQATOKTENLTPEDEVSKREGINAEOIVIKITDOGYVT 66
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 120
Db 67 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 126
Qy 121 AHADNVRTKEEINROKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 180
Db 127 AHADNVRTKEEINROKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 186
Qy 181 GDAYIVPHGDHYHYPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 240
Db 187 GDAYIVPHGDHYHYPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 246
Qy 241 VSNPGTTNTNTSNNSNTNSQASQNSNDISLKLQYKPLPSQRHVESDGLVFPDPAQITST 300
Db 247 VSNPGTTNTNTSNNSNTNSQASQNSNDISLKLQYKPLPSQRHVESDGLVFPDPAQITST 306
Qy 301 ARGVAVPHGDHYHYPYSOMSELEERIAIIPLRYSRSHMWVDSRPEOPSPQPTPEPSPG 360
Db 307 ARGVAVPHGDHYHYPYSOMSELEERIAIIPLRYSRSHMWVDSRPEOPSPQPTPEPSPG 366
Qy 361 POPAENLKIDSNSLVSQLVRKVGEGYVEEKGISRYVFAKDLPSETYKNLESKLSKQES 420
Db 367 POPAENLKIDSNSLVSQLVRKVGEGYVEEKGISRYVFAKDLPSETYKNLESKLSKQES 426
Qy 421 VSHITLAKKENVAPRDOEFYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 480
Db 427 VSHITLAKKENVAPRDOEFYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 486
Qy 481 KEKLVDDLAFAPITTHPERLGPNSQIETYTEDEVRIADLADKYTTSDDGYIFDEHDIISD 540
Db 487 KEKLVDDLAFAPITTHPERLGPNSQIETYTEDEVRIADLADKYTTSDDGYIFDEHDIISD 546
Qy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATYKKGILPPSPDADVKANPTGDSAAAIY 600
Db 547 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATYKKGILPPSPDADVKANPTGDSAAAIY 606
Qy 601 NRKGEKRIPLVRLPYMEHVEYEVKNGNLIIPKHDYHNIRKFAWDDHTYKAPNGYTTLED 660
Db 607 NRKGEKRIPLVRLPYMEHVEYEVKNGNLIIPKHDYHNIRKFAWDDHTYKAPNGYTTLED 666
Qy 661 LEATIKYVEHDERPHSDGNGNASEHVLGKDHSEDPNKFKADEPVEETPAPEVP 720
Db 667 LEATIKYVEHDERPHSDGNGNASEHVLGKDHSEDPNKFKADEPVEETPAPEVP 726
Qy 721 QVETKEVEAQLKEAEVLLAKYVDSLSKANATETLAGLRNNLTLOIMDNNSIIMAEKELLA 780
Db 727 QVETKEVEAQLKEAEVLLAKYVDSLSKANATETLAGLRNNLTLOIMDNNSIIMAEKELLA 786
Qy 781 LKGSNPSVSKEKIN 796
Db 787 LKGSNPSVSKEKIN 802

```

RESULT 2  
09AHT9 PRELIMINARY: PRT: 816 AA.

```

ID 09AHT9 AC 09AHT9: 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_taxid=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4.
RX MEDLINE=21116976; PubMed=11179332;
Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

```

```

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
RT Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL; AF291695; AAK19155.1; -
SQ SEQUENCE 816 AA; 91519 MW; 5359126A61D27ED CRC64;

```

Query Match 100.0%; Score 4163; DB 2; Length 816;  
Best Local Similarity 100.0%; Pred. No. 2,1e-222;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 SYELGLYQARTYKENVNSYIDGKQATOKTENLTPEDEVSKREGINAEOIVIKITDOGYVT 60
Db 21 SYELGLYQARTYKENVNSYIDGKQATOKTENLTPEDEVSKREGINAEOIVIKITDOGYVT 80
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 140
Qy 121 AHADNVRTKEEINROKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 180
Db 141 AHADNVRTKEEINROKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 200
Qy 181 GDAYIVPHGDHYHYPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 240
Db 201 GDAYIVPHGDHYHYPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 260
Qy 241 VSNPGTTNTNTSNNSNTNSQASQNSNDISLKLQYKPLPSQRHVESDGLVFPDPAQITST 300
Db 261 VSNPGTTNTNTSNNSNTNSQASQNSNDISLKLQYKPLPSQRHVESDGLVFPDPAQITST 320
Qy 301 ARGVAVPHGDHYHYPYSOMSELEERIAIIPLRYSRSHMWVDSRPEOPSPQPTPEPSPG 360
Db 321 ARGVAVPHGDHYHYPYSOMSELEERIAIIPLRYSRSHMWVDSRPEOPSPQPTPEPSPG 380
Qy 361 POPAENLKIDSNSLVSQLVRKVGEGYVEEKGISRYVFAKDLPSETYKNLESKLSKQES 420
Db 381 POPAENLKIDSNSLVSQLVRKVGEGYVEEKGISRYVFAKDLPSETYKNLESKLSKQES 440
Qy 421 VSHITLAKKENVAPRDOEFYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 480
Db 441 VSHITLAKKENVAPRDOEFYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 500
Qy 481 KEKLVDDLAFAPITTHPERLGPNSQIETYTEDEVRIADLADKYTTSDDGYIFDEHDIISD 540
Db 501 KEKLVDDLAFAPITTHPERLGPNSQIETYTEDEVRIADLADKYTTSDDGYIFDEHDIISD 560
Qy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATYKKGILPPSPDADVKANPTGDSAAAIY 600
Db 561 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATYKKGILPPSPDADVKANPTGDSAAAIY 620
Qy 601 NRKGEKRIPLVRLPYMEHVEYEVKNGNLIIPKHDYHNIRKFAWDDHTYKAPNGYTTLED 660
Db 621 NRKGEKRIPLVRLPYMEHVEYEVKNGNLIIPKHDYHNIRKFAWDDHTYKAPNGYTTLED 680
Qy 661 LEATIKYVEHDERPHSDGNGNASEHVLGKDHSEDPNKFKADEPVEETPAPEVP 720
Db 681 LEATIKYVEHDERPHSDGNGNASEHVLGKDHSEDPNKFKADEPVEETPAPEVP 740
Qy 721 QVETKEVEAQLKEAEVLLAKYVDSLSKANATETLAGLRNNLTLOIMDNNSIIMAEKELLA 780
Db 741 QVETKEVEAQLKEAEVLLAKYVDSLSKANATETLAGLRNNLTLOIMDNNSIIMAEKELLA 800
Qy 781 LKGSNPSVSKEKIN 796
Db 801 LKGSNPSVSKEKIN 816

```

RESULT 3  
09AG74 PRELIMINARY: PRT: 844 AA.

AC 09AG74; 2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
DE PkPA.  
GN PkPA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP  
RP  
RP MEDLINE=21246685; PubMed=11349048;  
RX Zhang Y., Masi A.W., Barniak V., Mountzouros K., Hostetter M.K.,  
RA Green B.A.; "Recombinant PkPA Protein, a Unique Histidine Motif-Containing Protein  
RT from Streptococcus pneumoniae, Protects Mice against Intranasal  
RT Pneumococcal Challenge.";  
RL Infect. Immun. 69:3827-3836(2001).  
SQ EMBL, AF340221; AKK26629.1; "-"  
SQ SEQUENCE 844 AA; 94769 MW; D738A5520FF8902 CRC64;

Query Match	77.7%;	Score 3236;	DB 2;	Length 844;
Best Local Similarity	75.3%;	Pred. No. 4.6e-171;		
Matches 620;	Conservative 66;	Mismatches 99;	Indels 38;	Gaps 4;

QY	1	SYELGLVAFVKKENNRYSYIDGKAOTQKTEMLPDEVSKREGJNABOIVIKITDQGYVT	60
Db	21	SYELGLVAFVKKENNRYSYIDGKAOTKTEMLPDEVSKREGJNABOIVIKITDQGYVT	80
QY	61	SHGDHYHYNKGKVPDIAITISELLMKDDPNYKIAKEDIINENVAGGYIVYDKGRYYVYLDA	120
Db	81	SHGDHYHYNKGKVPDIAITISELLMKDDPNYQIKDEDIITISEIKGGYIVIKYDQKYYVYLDA	140
QY	121	AHADNVRKEELINQKQSHSQHREGGTPRNGGAVALASQGGYTTDDGYIFPAASDIEDT	180
Db	141	AHADNVRKEELINQKQSHSQHREGGTPRNGGAVALASQGGYTTDDGYIFPAASDIEDT	200
QY	181	GDAYVPHGDHYHYTPKRELASAEALAAEAFSGRGJNSNSTRYRONSDNSTRNWPVS	240
Db	201	GDAYVPHGDHYHYTPKRELASAEALAAKAPLSGNGJNSNSTRYRONSDNSTRNWPVS	260
QY	241	VSNPCTTNTSNNSNTNSQASQSNDDIDSLKOLYKPLPSORHVESDGLVEDPAOITSRT	300
Db	261	VSNPCTTNTSNNSNTNSQASQSNDDIDSLKOLYKPLPSORHVESDGLVEDPAOITSRT	320
QY	301	ARGVAVPHGDHYHYTFPYQSOMSELEERIRIITPLRKSNNHWVDSRPEOPSPOTPEPSG	360
Db	321	ANGVAVPHGDHYHYTFPYQSLPTEKLARIITPLRKSNNHWVDSRPEOPSPOTPEPS	380
QY	361	POPAPNLT - IDSN -- SSLVSQLVRRKVGEGYFEEKGISRYVEAANDLPSETVKNLESKLS	416
Db	381	POPAPNPAPASNPIDEKLYKAEVARKVYDGYVEENGVPRTIPADLSAETAGDISKLA	440
QY	417	KOESYSHLTFFKKEVVARDOEFYKAAVNLTEAKALFXNKGSRNSDQALDKLEERLND	476
Db	441	KOESYSHLTGAKKTDLPSSDRFTYKKAIDLARITHODLDNKGROYDEALDNLERLKD	500
QY	477	ESTNKEKLVDDLFLABITPHERLKGPNSSOILETDEVEYRIAQLADKYYTSDGYIFDEHD	536
Db	501	VSSDGVKLVDDLFLABITPHERLKGPNNAQITTYDDEIQVAKLAGKYTTEDGYIFEDPD	560
QY	537	IISDGGAAVYVPHMGSHWIGKDSIDSEKEXAAQAAYTEKGLLPSPDPADVANKPTGDSA	596
Db	561	IISDGGAAVYVPHMTHSHWIKKDSISEEARAQAAYAEKGLTPSPDTHODSGNTEARGA	620
QY	597	AAIYNRVGEKRIPLVLRPYWEHTEVANKGNLITLPHKDHYINIKFAFEDDHTYKAPNGY	656
Db	621	EAIYNRVAAAKVPLDRBPNYQIYVEVKNCSLIIPIRHDYINIKFEWDEBELYAPKCY	680
QY	657	TLEDLFATIKYIVHEPDRPHSNDGMASEHYLKGKQHSDEPNKFKRADEE-----PV	710
Db	681	TLEDLFATIKYIVHEPDRPHSDNDFGNASDHVQENKNGQADTNPOTEPKNEPOTEPKE	740

```

QY 711 EET-----PAPPEVQVETEVEEQLAEAEVLAKT 742
    | | | | | | | | | | | | | | | | | | | |
Db 741 EETREKPKQSEKPEKPTPEPEEESPEESPEEQVETEKVKRLAEADLLKIQ 800
    | | | | | | | | | | | | | | | | | | | |
QY 743 DSSLKANATEFLAGLRNNLLIQIDNNNSIMAEAEKLLALLKGS 785
    | | | | | | | | | | | | | | | | | | | |
Db 801 NPITKSNAREFLTGKIKNNLLFGTQDNNNTIMAEAEKLLALLKES 843
    | | | | | | | | | | | | | | | | | | | |

```

RESULT 4	
Q97QM9	
ID Q97QM9	PRELIMINARY; PRT; 819 AA.

DT 01-OCT-2001 (TREMBLrel . 18, Created)  
 DT 01-OCT-2001 (TREMBLrel . 18, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel . 19, Last annotation update)  
 DE Conserved domain protein.  
 GN SP117A.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales  
 OC Streptococcaceae; Streptococcus.  
 NCBI\_TaxID=1313;

RP SEQUENCE FROM N.A.  
 RT  
 KC STRAIN-TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Debby R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Rauden D.,  
 RA Holtzapfe E., Khouri T.W., Wolf M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.E., Feilbuhl Y.T.V., Anguoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollinshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*  
 pneumoniae".  
 RL Science 293:498-506(2001).  
 DR EMBL; AE007418; AAK75283.1; -.  
 DR TIGR; SP1174; -.  
 KW Complete proteome.  
 QO SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match	67.2%;	Score 2797.5;	DB 16;	Length 819;
Best Local Similarity	67.0%;	Pred. No. 8.4e-147;		
Matches 546;	Conservative 84;	Mismatches 138;	Indels 47;	Gaps 7;

QY	1	SYGLGLVOA - RTVENNRYVSJIDKQAKTQKTEMLTPBEVSKREJINAEQIVIKITDQGY	53
Db	21	SYELGRIOAGDQKESNNVAIITSDQAGQKRENTLPBEVSKREJINAEQIVIKITDQGY	80
QY	60	TSHSDHHYHYNGKAPYPAIISEELMLKDPNKALEDEDIVNEVKGGYIKYDQKYYVLK	119
Db	81	TSHGHHYHYNGKAPYDAIISEELMKDPNQTKLSDIVNEIKGGYIKYNGKYYVLK	140
QY	120	AAHADNVTKEINROKQEHSHQHEGGTPRFDGAVALARSGRTTDDGXIIPNADIIED	179
Db	141	AAHADNITKEIEIKROKQERSHINNS -- RADNVAARAAQGRATTDDGXIIPNADIIED	197
QY	180	TGDAYIVPHGHYHYIPKNELSASELAEEAFISGRCLNSMTRYBRONSDNTSRTWMP	239
Db	198	TGDAYIVPHGHYHYIPKNELSASELAEEAFYNGS -----KQGRPSSSSSYNA	246
QY	240	SYSNEGTTNTNNSNNTNSQASNDIDSLKOLYKPLSQRHVESDGLVDFPAQITSR	299
Db	247	NPAPRPLSENNLITVPTTYH-H-NGENISILSTRILYAKPLSERHVESDGLIFDPAQITSR	305
QY	300	TARGAAYPHGHYHYIPYSOMSELEENRIARIIPRYSNMHWPPSRDEQSPQTPPEPSP	359
Db	306	TARGAAYPHGHYHYIPYEQMSLEKRIARIIPRYSNMHWPPSRDEEESPQCTPEPSP	365
QY	360	GPOEAPNLKIDSNSSLVSOLVRRKQGEVVEEKGISRYVEAKDLPESTYKNLSSLSKOE	419

Db	366	SPQAPSNPID--EKIVKEAVRKVGDGVYEENSVSYIPAKOLSMETAGIDISKLAKOE	423
QY	420	SVSHTLTAACKENVAPRDOEYDRAVNLITEAHKALEYXKGRNSDFOALDKLERLND	479
Db	424	SLSHKLTGKTKTDLPSSDSREFYKRAVYDLARIHODLNKRGROVDFEALDNLIERLDV	483
QY	480	NKEKIVDLDLAFAPITHPBEIGKPNISOITEBEVAVIQAOLAKYTTSDGYIFEDHDI	539
Db	484	DKVATVEDLAFAPIRIRPERLGPAPNOITYTDDIYVATLAKGYTTEDEGYITDPRI	543
QY	540	DEGDAAVYPRHAGSHWIGKSLDSKEKEVAAQVAKKEGILPPSPBDADVKNPMPGDS	599
Db	544	DEGDAVYVPRHHTSHMWIKKSLDSAEBAQAQAVAKKEGLPPSTHDQSDNTAKAG	603
QY	600	YNRKVGKRIPLVLRPIVVEHTVYVUKGNLIIPIKDHVNIKRAMFDDHTYKAPNG	659
Db	604	YNRKAAKRVKLDMPRPNYLQTYVEVKNGSILIPHYDHNHIKEWFEGLYEAPKGT	663
QY	660	DLFATIKYVEHPDERPNSNGWENASEHYLGGKDHSDSENNKRNKAEE-----	713
Db	664	DLATVAKYVEHPDERPHSDNGFGNABDHYORKNNGADLNQTEKPESEKPYOTEK	723
QY	714	-----PAEPVPOVETEKEVLAQLEAEVLLARVDSILKANA	750
Db	724	PREKPOSEKPESPKPTPEEPESEPEEPOVETEKEVEKTLREAEIDLGIOPITK	783
QY	751	TEITLAGLRNNITLQIMONNSIMAAEKLLALLKKS	785
Db	784	KETLTGLKNNLITFGTDNNITIMAAEKLALLKKS	818

RESULT 5  
Q9ANY3  
ID Q9ANY3 PRELIMINARY; PRT; 819 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE Pnuemococcal histidine triad protein B precursor (Fragment)  
NC PHTB.

0C Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
0C Streptococcaceae; Streptococcus.  
0X NCBI\_TaxID=1313;

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=21101045; Pubmed=11159990;

RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,  
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,  
RA Langermann S., Koenig S., Johnson S.;

RT "Identification and characterization of a novel family of pneumococcal  
RT proteins (the Pht family) that are protective against sepsis." ;  
RL Infect. Immun. 69:949-958(2001).

DR	EMBL; AF318954; AAK06759.1; -.
KW	Signal.
FT	SIGNAL 1 29 POTENTIAL.

FT	NON_TER	819	819
SQ	SEQUENCE	819 AA;	92108 MW; E602CFC16CC28A5F CRC64;

Query Match	66.9%	Score 2786.5;	DB 2;	Length 819;
Best Local Similarity	66.7%;	Pred. No. 3.4e-146;		
Matches 544;	Conservative 84;	Mismatches 140;	Indels 47;	Gaps 7

QY 1 SYELGLYQ-RTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAQIYIKITDQGV 59

```

21 SYELGRYQAGODKKESNRVAYIDGQAGQKAENLTPDEVSKREGINAQIVIKITDQGYV 80

```

07 TSHGDH IYNGKVPDAI ISEELIMDPN KKLDEDEIVN EVRKGGYX IAKVDKXIVY LKD IIS

05 120 AAHADNVTKKEEINROKOEHSOHEGTPBNDGAVALARSOGRYTDDGYIFNADIIED 179  
 22 01 IONQDIIHIIINGNAVEIDALLTSEBDEHNFENIYENDDDIYMLNGSLVAVNGNIIYIIND 178

[illegible][illegible][illegible]

DB 24 / NPAQPKSENNHNLIVIPRIHQ - NQGENISSLLKRELIANPUSKEHVESDGLIFDPAQISK 303

OY 300 TARGVAVPHGDHYHFIPIYSONSELEERTARIIPILYRSNMWVPDSRPQSPQPTPEPSP 359

[illegible]

```
07 GFQFRLNLDSCSDVSSQLVAVGVGVVELENGSNTVFANVDZSLVANNEDDSDKVL  
    ||||| : |||||:||||| :|||: ||| :|||: |||  
Db 366 SPQPAPSPID--GKLTKAEAVRKVGDDGVFEENGYSRYIPAKDLSELAAGIDSRLAQE 423
```

QY 420 SVSHTLTAKENVARPDQEFDKAYNLLTEAHKALFXNKGRNSDFQALDKLLERINDEST 479  
| : | | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 424 SLSHLTGMKTDPSSDREFNKAYDLARIHODLIDNKGVOVFEALDNLLERIKDVSS 483

QY 480 NKEKLVDDLAFAPITTHPERLGKPNISOIYEYTEDEVRIAOLADKYTSDGYIFDEHDIIS 539  
:  
:

Db 484 DKVKLVEDILAFAPIRHPERLGKPNQAITYTDEIQVAKIAGKYTAEDGYIFPRDITS 543

Ov 540 DEGDAYVTPHMGSHWIGKDSIDSEKYAAQAAVYKEKGIPEPPSPDADVYKANPTGSAAI 599

```
Db      544 DEGDAYTPHMTSHWIKKDSLSAERAAQAYAXEKGITPPSTDHQDSCNTEAKGAEI 603
```

Qy 600 YNRVKGGEKRIPLVNLPLYMEITVEVKNGNLIIPKHDYHNHNTFAWFDHITYKAPNGTYL 659  
 |||| :::: || :::: ||||| :||| :||| ||| :||| ||||  
 Db 604 YNRVKAAKKAVPLDMRPYNLQYTVVEVKNSLIIPHDYHNHNTFEWFDCLYEAPRGTYL 663

660 DLFATIKYVEHPDERPHSNDGWSASEHYLGKDKHSEDPNKNFKADE-----PVEET 713

Db 664 DLTATVYVEHPRPHSDNGFGNASHDVQRNKGADTNOTKPSSEKPTKEPPEET 723

QY 714 -----PAPEVPQVETEKEVEAQLKEAEVLAKVTSSLKANA 750

```
Db      724 PREEKPOSEKPTEEPESPESEBEPQVEIKVEEKLREADDLGKIQDPITIKSNA 783
```

QY    751 TETLAGLRNNLTQIMDNNSIMAEAKLALLKGS 785  
      ||| :|||     ||| :||||||| |  
DB    784 KETLTGLKNLLFGTDNNTIMAEAKLALLKES 818

## RESULT 6

Q9ANY2	PRELIMINARY;	PRT;	839 AA.
ID Q9ANY2			
AC Q9ANY2;			

DT 01-JUN-2001 (TREMblurel, 17, Created)  
DT 01-JUN-2001 (TREMblurel, 17, Last sequence update)  
DT 01-OCT-2001 (TREMblurel, 18, Last annotation update)  
DT 01-DEC-2001 (TREMblurel, 19, Last annotation update)  
DT 01-DEC-2001 (TREMblurel, 19, Last annotation update)

DE Pneumococcal histidine rich protein D precursor (Hypometrical prote  
DE SPI003) (Fragment).  
GN PHTD OR SPI003.

0C Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
0C Streptococcaceae; Streptococcus.

OX	NCBI_taxid=1513;
RN	[1]
RP	SEQUENCE FROM N.A.
RY	MEETING=3110105. Pubmed=11150000.

RA MEDLINE=21101045; PubMed=11139390;  
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,  
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,  
RA Lutzmann S., Kocic S., Johnson S.,

RT "Identification and characterization of a novel family of pneumococcal  
RT proteins (the Pht family) that are protective against sepsis.";  
RT Infect Immun 69:949-958(2001)

RE INTEL. Intern. 03.949 900(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRATN=TIGRA.



```
QY 241 VSNPGTTNTNTSNSTNSQASOSNDISLKLQLYKLP.LSQRHVESDGLVDPDAQTST 300
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 -----NTOQVAKGSTSKRPANKSEMIQSLKELKLYDSPSAQRSESDGLVDPDAKIISRT 301
QY 301 ARGVAVPHGDHFTIPYSOMSELEERLARIPIRLYRSNMWVDSRREGSPQTPPEPSG 360
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 PNGVALPHGDHFTIPYSKLSALEEKIARWP----- 333
QY 361 POPAPMLKIDNSNSLVSQLVRKVGEGYVEEKGISRYVEPAKDLPSFTVKNLESKLSKQES 420
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 334 -----ISGTSYSTNAK-----PNEVSSLSGSLSSNPS 363
QY 421 VSHHTLAKENAVAPRDOEYDKAYNLLTEAHKALEXNKGNSDFQALDKLERLNESTIN 480
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 364 ----LTTSKELSSASDGYIFENPK-DIVEETAAIYIVRHG--DHFHYIPK-----SNOIG 410
QY 481 KKLVDLLAFLAPITHPERLKPNSQIETDEVRILQALADKYTSDQYTFDEHIIISD 540
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 411 QPTLPNNSLATPSP-SLPINPGSHEKHE-----EDGYGFDPANRIIE 452
QY 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKEGILPPSPDADYKANPTGDSAAIY 600
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 453 DSGFVMSGHDHNYEFKKDLTEEOJKAQKHLE-----VKTSNGDLSLSH 501
QY 601 NR-----VKEKRIPLYRLPYMV--EHTVEYKNGNLI-PRKDHYNIKFAM 644
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 502 EODYSNAKEMKDLDKIEKIAIGKQYKRESIVYKKEKNAIIYPPGHDDHDP-- 558
QY 645 FDDHTYKARNGYTLBEDFTIKYVEHPDERPHSNDG-----WGNASEHVLGKKHSDEN 700
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 559 IDEH--KRVGIG---HSHSNTELFKPEEGVAKKEGKNVYTGEBELTNVNLKSTEN 611
QY 701 KNF 703
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 612 ONF 614
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
Q99XV4 PRELIMINARY; PRT; 825 AA.
AC Q99XV4;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE Hypothetical protein SPY2006.
GN SPY2006.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1.
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Perleaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RU Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AB006623; AAK34688.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF509696D50F4 CRC64;

Query Match 22.8%; Score 949; DB 16; Length 825;
Best Local Similarity 29.5%; Pred. No. 1.9e-44;
Matches 277; Conservative 100; Mismatches 214; Indels 348; Gaps 34;

QY 1 SYELGIYQARTYKENVRSYID---GKOATOKTEMLTPDEVSKREGINAEOIVIKITDGG 57
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 22 SYQLGKHNGMSATKQNGIAYIDDSKSKAKAPRT-NKTMQJIAEBESIMQIVKTKITDGG 80
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 58 YVTSHGDIHYNYNGKVPYDAIISSELLMKDPMYKCLKDEDIVNEVGKGYIVKVDGKIYVL 117
```

```
Db 81 YVTSHGDIHYNGKVPYDAIISSELLMTDPNVRKQSDVINELIDGYIKVNGVYVL 140
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 118 KDAAHADNVRTEKEINROKQESOH-REGG-----FPRNGAVALARSGRYTTDDCY 169
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 141 KFGSKRKNIARTQOIALEOVAKGTKEAKEGLQVAHLSKEEYAAVAENAKRQGRITTTDDCY 200
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 170 IFNASDIIDTDGDAYIVPHGDHYHYIPKNLSASELAEEALS--GRGNLSNRYRR 226
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 IFSPPTIIDLDGDAYLVPHGNHYHYIPKDLSPSELAQAQVAWSQKQGRG--ARPSYR 258
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 227 QNSDNTSRTNM--VPYSV-NPGTTNTNTSN-----NSNTNSQASQSDN-----IDSL 271
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 TPAPAPGRKAPIPDVTAPNGQHQPNDNGYHPAPRPDASQNRHQRDEFGKTFKELL 318
QY 272 KQLYKLP.LSQRHVESDGLVEDPAQTSRTARGAVAPHGCHYFTIPYSOMSELEERLARI 331
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 DQHLRLDLTRYVEEDGLTFEPTQYIKSNAGFYVYPHGDIHIIITRSQSLPEMELAD-- 376
QY 332 PLRYRSNMWVPSRREGQSPQTPPEPSGPOPAPMLKIDNSNSLVSQLVRKVGEGYVEE 391
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 377 --RYLAGQTEDD-----DSGSD----- 391
QY 392 KGISRYVFAKDLPSFTVKNLESKLSKQESVSHTLAKKENAVAPROEYDKAYNLLTEAH 451
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 392 -----HSPKPSDKE-VTHTFLGHR-----IKAY----- 412
QY 452 KALFNKNGNSDFQALDKLERLNESTNKEKLVDDLLAFLAPITHPERLKPNSQIET 511
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 413 -----GKGLD-----GNP----- 420
QY 512 EDEVRLQADRYTTSQDGYIFDEHDIISDEGDAYVT PHMGSHWIGKDSLSDKEKVAQA 571
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 -----YDTSDAYVFSKESIHSYDKSGYATAKHGDHFHYIGFELDQYELDEVAN 468
QY 572 YTKKEGILPPSPDADYKANPTGDSAAIYNRKYGKRIPLY-----RLPYV 618
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 469 WVKANG-----QADELAALDQDQGEK-PLFDYKVKVSRKYTKDGKGYMM 513
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 619 -----TVEYKN-----GN 628
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 514 PKDGKDYFARQDLDTQAFABQELMLKDKKHRYRDIYDTGIERLAVDSSLPMHAGN 573
QY 629 -----LIIPKDHYNIKFAWEDHTYKAPNGYTLDELFTATIKYVEHPDERPHSND 680
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 574 ATYDGSFVIPIDHIDHIVYSWL-----TRDQJATIKYVMQHPREVR--D 618
QY 681 GW-----GNASEHYLGKDHSDPNKFKADEPVEETPAE-----PE----- 718
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 619 IWSKPGHEESGVIPNVPYPLDKRAGMPNMQIISAEYQKALAEGRFATPDGYIFEDPDY 678
QY 719 -----VPOVE-----TEKVEAOLKEAVLLAKVTDSLSKANATETLA 755
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 679 LAKETFWKDGFSFIPRADGSSLRTIINKSDLSQAEQQAQELLAAKN---AGDATIT-- 732
QY 756 GLRNMLTQIMDNNSIMEAEKLLALLKGSNPPSSVSKER 794
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 733 -----DKPREKQOADK-----SNEHQPSSEASKEE 757
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
Q93GT5 PRELIMINARY; PRT; 825 AA.
AC Q93GT5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
```

[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SI-9;  
RA Terao Y., Kawabata S., Hamada S.;  
RT "Characterization of a novel histidine triad protein of group A streptococci";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB073859; BAB1774.1; -  
SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFBF CRC64;

Query Match 22.7%; Score 944; DB 2; Length 825;  
Best Local Similarity 29.4%; Pred. No. 3.3e-44;  
Matches 276; Conservative 101; Mismatches 214; Indels 348; Gaps 34;

OY 1 SYELGLYQARTYKENVKRRVSYID--GKQATOKTENLTPDEVSKREGINAEOIVIKITDOG 57  
DB 22 SYQLGKHHMGSLTKDNQIAYIDDSKGAAPRT-NKIMDQISAEGISAEOIVAVKITDOG 80  
OY 58 VYTSHGDIHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVGKGYIVKDGKYYVL 117  
DB 81 VYTSHGDIHYHYNGKVPYDAIISELLMTDPYRKQSDVINELIDGYIVKNGVYYVL 140  
OY 118 KDAADADVNRKEELNRKQKESQH-REGG-----TPRNDGAVALARSGRYTTDDGY 169  
DB 141 KPGSRKKNIKRTKOQIAEOVAKGTKEAKKGLAOVAHLSKEEVAAYNEAKRQGRYTTDDGY 200  
OY 170 IFNADIIEDTGDAIVPHGDHYHYIPKNELSASELAAEAFLS--GRNLNSRRTYR 226  
DB 201 IFSPDIIIDDLGDAIVLPHGNHYHYIPKDLSPSELAAQAQVWSQKQGRG--ARPSDYR 258  
OY 227 QNSDNTSRTNW-VPSVS-NPGTNTNTSN-----NSNTSQASQSDN-----IDSL 271  
DB 259 TPAPARRKAPIPDVTNPGQHQPNQNGYHAPARRPNDAQONKHQDEFKQKFKELL 318  
OY 272 QLYKLPESQHVESDGLVFPDAQITSTARGAVAPHGDIHYIPYQMSLEERIALRI 331  
DB 319 QLRDLKTRYHVEDGLTFEPTQYKSNAGYVYPHGDIHYIPRSQISPLEMELAD-- 376  
OY 332 PLKRSNMHWVPSRPEQSPQPTPEPSGPQAPAPLKLIDNSLSQLVRKVEGYVEE 391  
DB 377 --RYLAGQTEDD-----DQSGD----- 391  
OY 392 KGISRYVEAKDLPESETVKNLESKLSQESVSHTLAKKENVAPRODFYDRAVNLITGAH 451  
DB 392 -----HSKPSDKE-VTHFTLGR-----IKAY----- 412  
OY 452 KALFYKGRNSDFQALDKLERLNDESTNKEKLVDDLAFAPLITHPERLGRPSQIEYT 511  
DB 413 -----GKGLD-----GKP----- 420  
OY 512 EDEVRIAGLADKYTTSDGYIFDEHDIISDEGDVAVTPHMGSHWIGKDSLQKEKVAQA 571  
DB 421 -----YTSDAVYFSKESISHVDSGYAKAGDHFIHIGFGELEGYELDEVAN 468  
OY 572 YTKKGLLPSPSDADVANKPTGDSAAIYNNRYKGEKRIPLV-----RLPYAV 618  
DB 469 WVKAKG-----QADELAALDQEQKEK-PLFDYTKVSKRYTKDKGVGMM 513  
OY 619 -----EH-----TVEYKN-----GN 628  
DB 514 PKDGKDYFARQDLDLQIAFAEOELMKDKKHVRYDVIDTGIEPRLLAVDSSLEPMHAGN 573  
OY 629 -----LIPKHDYHNINIKFAWFDHTYKAPNGYTLIEDLFAITKKYVVEHDEPHAND 680  
DB 574 ATYDTGSSFVIRPHIDHIVPRYSL-----TRDQIATIKYWMQHEVAP--D 618  
OY 661 GW-----GNASEHVLGKRKDHSEDPKNKFKADEPEVEETPAE-----PE----- 718  
DB 619 IWSKPGHESSGVIPNVPPLDKRAGMPNMQIISAEVQKALABERAFATPGYIIPDRDV 678  
OY 719 -----VPOYE-----TEKVEAQLKEAEVLLAKVYDSSILKANATEPLA 755  
DB 679 LAKETFWKDGSSFSIPRADGSSLRTINKSDLSQAEWQAOEILAKKN-----AGDATDT-- 732

OY 756 GLRNNLTQIMDNNSIMAEKLLALKGSNPSYSKER 794  
DB 733 -----DKPKKQQAQAK-----SNENQDPSSEAKSKEE 757

## RESULT 10

O9ZHG7 PRELIMINARY; PRT: 822 AA.  
AC O9ZHG7;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
OS Hypothetical 92.4 kDa protein.  
OC Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1311;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R268;  
RX MEDLINE=99115568; PubMed=9916102;  
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heynemann J.,  
RA Schmitzler N., Luetticken R., Podbielski A.;  
RT "Lmb, a protein with similarities to the Irai adhesin family, mediates  
RT attachment of Streptococcus agalactiae to human laminin.";  
RL Infect. Immun. 67:871-878(1999).  
DR EMBL; AF062533; AAD13797.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 22.4%; Score 934; DB 2; Length 822;  
Best Local Similarity 29.4%; Pred. No. 1.3e-43;  
Matches 271; Conservative 99; Mismatches 235; Indels 318; Gaps 32;

OY 1 SYELGLYQARTYKENVKRRVSYID--GKQATOKTENLTPDEVSKREGINAEOIVIKITDOG 57  
DB 22 SYQLGKHHMGSLTKDNQIAYIDDSKGAAPRT-NKIMDQISAEGISAEOIVAVKITDOG 80  
OY 58 VYTSHGDIHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVGKGYIVKDGKYYVL 117  
DB 81 VYTSHGDIHYHYNGKVPYDAIISELLMTDPYRKQSDVINELIDGYIVKNGVYYVL 140  
OY 118 KDAADADVNRKEELNRKQKESQH-REGG-----TPRNDGAVALARSGRYTTDDGY 169  
DB 141 KPGSRKKNIKRTKOQIAEOVAKGTKEAKKGLAOVAHLSKEEVAAYNEAKRQGRYTTDDGY 200  
OY 170 IFNADIIEDTGDAIVPHGDHYHYIPKNELSASELAAEAFLS--GRNLNSRRTYR 226  
DB 201 IFSPDIIIDDLGDAIVLPHGNHYHYIPKDLSPSELAAQAQVWSQKQGRG--ARPSDYR 258  
OY 227 QNSDNTSRTNWVPSVS-NPGTNTNTSN-----NSNTSQASQSDN-----IDSL 272  
DB 259 TPAPARRKAP-IPDVTNPGQHQPNQNGYHAPARRPNDAQONKHQDEFKQKFKELL 317  
OY 272 QLYKLPESQHVESDGLVFPDAQITSTARGAVAPHGDIHYIPYQMSLEERIALRI 332  
DB 319 QLRDLKTRYHVEDGLTFEPTQYKSNAGYVYPHGDIHYIPRSQISPLEMELAD-- 374  
OY 332 PLKRSNMHWVPSRPEQSPQPTPEPSGPQAPAPLKLIDNSLSQLVRKVEGYVEE 391  
DB 377 --RYLAGQTEDD-----DQSGD----- 391  
OY 392 KGISRYVEAKDLPESETVKNLESKLSQESVSHTLAKKENVAPRODFYDRAVNLITGAH 452  
DB 392 -----HSKPSDKE-VTHFTLGR-----IKAY----- 410  
OY 452 KALFYKGRNSDFQALDKLERLNDESTNKEKLVDDLAFAPLITHPERLGRPSQIEYT 512  
DB 413 -----GKGLD-----GKP----- 418  
OY 512 EDEVRIAGLADKYTTSDGYIFDEHDIISDEGDVAVTPHMGSHWIGKDSLQKEKVAQA 572

```

Db 419 -----YDSDAYVFSKESISHVDKSGVTAKGHDHFHYIGFLEOYELDEVANW 467
QY 573 TREKG---ILPPSPDADYKAMPTGDSAAA-----LYNR----- 602
Db 468 YKAKGQADLVALLADDOEGKEKLEPDTKVKSRKYTKDGAVGYIMPKDGKDYIYARQOLD 527
QY 603 -----VKGEKR-----IPVRLPYMEHTEVEVANGNLIIPKHKH 636
Db 528 TQIAFAEDLMKKDKKHRYDYIDVTGIEPRLAVDVSSLMHMGNNATYDGTSSFPVPHIDH 587
QY 637 YNNIKFAWDDHTTKAPNCTYLEDLFATIKIYYEHDPDEPHSNDG-----GNASHE 688
Db 588 IHVVPYSWL-----TRNQATIKIYVQHPERB---DVMSKPGHESSGVIN 632
QY 689 VLGGKHDSDPNKFKADEPEYETPAE-----PE-----VP 720
Db 633 VTPLDKRAQMPNMQIHSABEYOKALAEGRFAAPDGYITDPDVLAKETFWKDKGSFSITP 692
QY 721 QVE-----TEKYAOLKEAEVLLAKVTDLSLKANATELTLAQRNNLTQIMDNNSI 771
Db 693 RADGSSLRTINKSDLSQAEMQOAOELLAKN---AGDATDT-----DKPBE 735
QY 772 MAEAEKLLALKGNSPSSYSKEK 794
Db 736 KQOAK---SNEQOQPSKSEE 755

```

## RESULT 11

```

Q9AE21 PRELIMINARY; PRT; 289 AA.
ID 09AE21
AC 09AE21;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Gränlund M., Michel F., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and GBS11, an active group
RT II intron identified in human isolates of group b streptococci.";
RL J. Bacteriol. 183:2560-2569(2001).
DR EMBL; AJ290952; CAC35985.1; -.
KW Hypothetical protein.
FT NON_TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;

```

Query Match . 16.3%; Score 679.5; DB 2: Length 289;  
 Best Local Similarity 53.8%; Pred. No. 3.6e-30;  
 Matches 140; Conservative 40; Mismatches 61; Indels 19; Gaps 8;

```

QY 1 SYELGLYQARTYKNNRNVYID---GKQATOKTENTLPDEVSREGINAEQIYIKITDGG 57
Db 22 SYQLGHNHGLTKDNQIAYIDSKGKAPKT-KNTMQISAEBISAEQIYIKITDGG 80
QY 58 VYTSHGHDHYHYNGKVPYDAIISELLMKDPNYKDKEDIVNEVGKGYIKYDGKYYVL 117
Db 81 VYTSHGHDHYHYNGKVPYDAIISELLMTDPNRFKQSDVINELIDGYIKVNGNYYVL 140
QY 118 KQAADADNVRTEELNRQOEHSQH-REGG-----TRNDGAVVALASQGRYTTDDGY 169
Db 141 KGSKRKNIRTKQALAEQYAKGTRKAKEGLAQVLAHLSKEEVAANVEAKRQGRYTTDDGY 200
QY 170 INASDIIEDTGAYIVPGDHYHYIPKNELASASELAAEAFLS---GRGNLSNRTYR 226
Db 201 ITSPDIIIDLDGDAYIVPGNHTHYIPKDKLSPSELAAQATWSQKQNG--ARPSDYRP 258
QY 227 QNSDNTSRTNWPFSVS-NPG 245

```

```

Db 259 TPAPGRKRAP-IDVYTPNPG 277

```

## RESULT 12

```

Q99Z76 PRELIMINARY; PRT; 792 AA.
ID 099Z76
AC 099Z76;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE Putative internalin A precursor.
GN INLA OR SPY1361.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najat F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006574; AAK34188.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR Pfam; PF00560; LRR; 4.
DR SMART; SM00370; LRR; 4.
KW Complete proteome.
SQ SEQUENCE 792 AA; 87458 MW; 9D5E32288485ACE0 CRC64;

```

Query Match . 5.4%; Score 223; DB 16: Length 792;  
 Best Local Similarity 19.5%; Pred. No. 0.00028;  
 Matches 168; Conservative 116; Mismatches 279; Indels 300; Gaps 42;

```

QY 1 SYELGLYQARTYKNNRNVYIDGKQATOKT-ENLNPDEVSKREG--INAEQIYIKITDGG 57
Db 30 YPIRTKQSRKMGTSKRIPIPKSKTKNTKHGVAGVDPPTDGGFLTKDSYILSKTDGG 89
QY 58 VYTSHGHDHY-----YNGKVPYDAIISELL-----MKDP-NYKLD 94
Db 90 IVVDHGHSHFLFYADLKGSPFEYLIPKGAFLAKRAVAQSAQSGSKVADPHNHYEFNP 149
QY 95 EDIVNEVGKGYIKYDKYVYLKDAADNVRTEELNRQOEHSQHREGGTPRNDGAY 154
Db 150 ADIVAEADALGYTVRHDHRYILKSS-----LSGQYQAOAKQVATRLPQTSLSV 198
QY 155 ALARSQG-----RYTDDGYIFNADSIIEDTGAYIVPGDHYHYIPKNELASASELAAEA 210
Db 199 STATAGTIGLHFPISDQGFNGGIGVTKDSILVDHGHLPISFADL----- 248
QY 211 FLSGRGNLSNRTYKRONSDNTRTNW--VPSVSNPQGTNTNTSNNSNTNSQASQNDID 268
Db 249 -----RQG-----GMAHVADQYDPAKAEKAPAEFHQTPELSEKEVQ 286
QY 269 SLIKQIYKLPISORHVESGLVDPQAQI---TSRTAKVAAPVPHGDHYHFIYSONSELE 324
Db 287 EKLAYL-----EKLIGIDPSTIKRVETQDGLGLEYPHHDHAYL---MISDIE 332
QY 325 ERLARIPIPRYSNMHWVPSRP-----EQLSPLRYSNMHWVPSRP-----EQLSPLRYS 356
Db 333 --TGKIDP---DPAHIEHARELEKHKVGMDTLRALGDEDEYILDIVRTHDAPTPPSWE 386
QY 357 PSPGPOP---APNLKIDSNS-----SLVSQLYRVKVGEGY----- 387
Db 387 KDPNNMKEWLAIVIKLDGSKRDPLOQRKLSLLPNL-ETLIGLGFPIPKDISPVLQFKLK 445
QY 388 --VFEKEGISRVFAKDP-----SETVYNLESKLSKQESVSHTLTAKEN---VA 433

```



```

Db 446 OLLMTKGVTDYRFLDNMFQLEGIDISQNNLKDI--SFLSKYKNL--TLVAADNGIEDIR 502
QY 434 PRDOEYDKAYULLTEFAHKAFLFXNKGNSDFQALDKL--LEKLUNDES-----TNKE 482
Db 503 PLQO-----LPMV-----KFLVLSNNKISDLSPLASLHOLQELHIDNNOITDLSVSKRE 552
QY 483 KL--VD-----DLTALFLAP-----ITHPERL--GKPN-----505
Db 553 SLTVYDLSSNADVDLATQAPKLETLMDVDFKSHLDPLKNNPNLSLSINRAQOSLEG 612
QY 506 -----SOIETEDEVRAQLADK-----YTTSDGY--TPEHDIIISDEGA 544
Db 613 IEASSVYVVEAEKQIKSLVLDKQGSJTLFLDVTGNQJLSTLEGVNNFTALDILSVSKNQ 672
QY 545 YTPPHMGH-----SHWIGKDSLSD--KEKYAAQAYTKREKGLIPSPDADVKAMP 591
Db 673 LNNVLSKRNKKTVTNIDISH--NNISLADKLNEQHIFPAIKN--PFAVYEGSVGNG 727
QY 592 TQDSAAAIYNRVKGKRIPLVRLPYWVEETVEVKNGLIIPKHDYHNIAKFAFPDHTYK 651
Db 728 TAEKRAAMATFKAKESQASESHDYNNHNTYDEEGHA-----HEHRDK-----DDHDI- 776
QY 652 APNGTLEDLFATIKYYVEHPDE 674
Db 777 -----EHEDE 781

```

## RESULT 13

```

Q963T1: PRELIMINARY: PRT; 1078 AA.
AC 0963T1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Glutamate-rich protein (Fragment).
GN GLORP.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21313688; PubMed=11420113;
RA "Thelsen M., Thomas A.W., Jepsen S.;
RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
RT glutamate-rich protein (GLORP) from Plasmodium reichenowi."
RL Mol. Biochem. Parasitol. 115:269-273(2001).
DR EMBL; AF356828; AAK40236.1; -.
FT NON_TER 1
FT NON_TER 1078
SQ SEQUENCE 1078 AA; 123581 MW; 8D388D88B223913C CRC64;

```

Query Match 4.7%; Score 197.5; DB 5; Length 1078;  
 Best Local Similarity 19.6%; Pred. No. 0.011;  
 Matches 164; Conservative 136; Mismatches 347; Indels 191; Gaps 37;

```

QY 7 YQARTVKKNNRVSYIDGKQATQKTENLTP---DEVSKREGINAOIYIKITDQGV--- 59
Db 372 HAVNNVLQENNNHNLHLEQEKANIESFEKKNIDSEIILPENVEKEIIVDVSPKHNHE 431
QY 60 -----TSHGDHYHYNGVYPDAITSEELMKDPYKDKEDIVNEVKGVIYKDGKY 114
Db 433 TLEQSTSESEHEAVSEKNAHETVEHEAVSQESNPEKADND--GNVSQSNNEINENEF 489
QY 115 VYLKAAH--ADNVRT-----KERI-----NRQKQEHSG--HRGGCIPPRNDGAVALAR 159
Db 490 VSEKSESEHPANESSLEEAHQEELIVPEQNNQGESKLVNDGCGFEAAHQEELIVPEQ 549
QY 160 QGRYTTDDGYIFNASDIEDTGDAYIVPHGDHYH---YIPKNELSASFLAAAEFLSGR 215
Db 550 NNQGESKLVNDGCGFEAAHQEELIVPEQNNQGESKLVNDGCGFEAAHHEMFSSFE 609
QY 216 GULSNSRYTRONSCNT-----SRTNWPVSYPGTTNTNTNNSNTNSQASQSDNDISL 270

```

```

Db 610 SN-SELHEHEVESEDSPEPEHEEVEVSESNPEASENESSIEBAHQEELIESQNDSE 668
QY 271 LKQLYKLPLSQNHVESDGLVFPDPAQ--ITSRTARGAVANPBGDHYHPLPSQMSLEERAR 329
Db 669 LNE-----NELVSEKSVSEPAHEVIVSEKSVSEP--AEHVEIVSEKSVSEPAHEVBS 720
QY 330 IIPLRYSNHWVPDSRPEQSPQTPPEPS---PGQOPAPNL-----KIDNSSLYVS--OLVR 381
Db 721 VSE--KSASE--PSEHVESVSEQSNNEPEKKGDPKSPKPEFEIEKVDPYVLDQITE 776
QY 382 KVGEVVEEEKGISRYVEKADLPSETVKNLESKLSQESVSHTLFAKKENVAAPRDOEYD 441
Db 777 -----PNEVDLPQNPQEPVEPSFVKIEKVPSEEN-----KQSDVPEVEE--- 816
QY 442 KAVNLLTEAHKALFANKGNSDFQALDKLELNDESTKKEKLVYDDLAFAPITHPERL 501
Db 817 -----KENVSVEVEKQNPQELFEI-----PLKRDDE--I 844
QY 502 GKPNQIETEDEVRLAOLADKYTTSDGYITPEHDIIISDEGDAYVTPPHGSHWIGKDSL 561
Db 845 EKIEPELEF--EDVHTEQDLEHKITVDPELIEVEI-----PSELHENEVAHPET 892
QY 562 SDKEVVAQAAYTKREKGLIPSPDAD--VKANPTGDSAAAIYNRVKGKRIPLVRLPYW 618
Db 893 VEIEEYF-----PEPNQNNPEPEINEDDKSAHQEIEVVEEILPEEDKNEKV 940
QY 619 EH-TVEVKNKGNLIIPKHDYHNIAKFAFPDHTYKAPNGTLEDLFATIKYYVEHDERPH 677
Db 941 EHEIVEVEE--ILPEENNEKV-----EHE-----IYEVEEILPE 973
QY 678 SNDGNGNASEH--VLGKDKHSDPEPNKFNKADEPVEETPAPEPPOVETEKVBAOLKEAV 736
Db 974 ENB--NEKVEHEIVEVEELPEEDKNEKVEHEIVEVEELPE--ENKNEKVEHEIVEVEE 1029
QY 737 LLAQYDSSLKANATETIAGLRNNLTLOITMNNNSIMAEEKLLALLKGSNPPSVSEK 794
Db 1030 ILPEITE-----IEVPSQTNNEINIE-----TIKPEEK-----NEVSVEEK 1068

```

## RESULT 14

```

ID 077033 PRELIMINARY: PRT; 1390 AA.
AC 077033;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRPA.
GN Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AX2;
RA Saito J., Adachi H., Sutouh K.;
RT "Dictyostelium TRPA homologous to yeast Sn6 is required for normal
RT growth and early development."
RL J. Biol. Chem. 0:0-0(1998).
DR EMBL; AB009080; BAA33143.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 9.
DR SMART; SM00028; TPR; 9.
SQ SEQUENCE 1390 AA; 160421 MW; BAAB926656002DE CRC64;

```

Query Match 4.6%; Score 192.5; DB 5; Length 1390;  
 Best Local Similarity 18.1%; Pred. No. 0.03;  
 Matches 147; Conservative 131; Mismatches 291; Indels 245; Gaps 35;

```

QY 15 NNRVSYIDGKQATQKTENLTPDEVSKREGINAOIYIKITDQGYTSGDHYHYNGVYP 74
Db 690 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 736
QY 75 YDAITSEELMKDPYKDKEDIVNEVKGVIYK-----DGKYYVYLKDAAHADNVRT 128

```

```
Db 737 -----DNVNSKNNNDVLDRRYKGLIEREKTSPNGGR-----:|||:
QY 129 KEINROKOEHSOHRGCGPR-----NDGAVALA-----RSQGRYTTDGG 168
Db 776 NRQ-NDSDSGDRDNRGDRSDRIOETRYREYNNNNNNNNNNNNNNNNNNNNNNNN 834
QY 169 YIFNASDIIEDTGDAYIVP---HGDIHYIIPKNELSASELAAEAFLSGRGLNSRTYR 225
Db 835 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 885
QY 226 RQNSDNTSRTNWPVS-----SNPGTTNTSNNSTNSQASQSDND-DLLKQL 274
Db 886 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 941
QY 275 YKLPISQHVESDGLVFD-PAQITRT-ARGVAVPHGDHYHFIPIYQMSLEIRIARIIP 332
Db 942 ALSPOSSQIKDRREIILDEESDINERSKTRSPSI-----VAAEAKRETVIY 988
QY 333 LRYRSHHWVDSRPEQSPQTPPEPSPGPAPAPNLKI-DSNSSLVSQLVKVGEGYVEE 391
Db 989 DKERS-----PPIIITEKPEKQVEKVTPEKESLVEKVDKE-----NEK 1027
QY 392 KGSRYVFAKDLSEYVKNLESLKQESVSHLTAKKENVAPRODEFYDKANLLTEAH 451
Db 1028 ESPSSSSSEKELEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 1078
QY 452 KALFXKGNHSDFOALDKLER-----LNDESTNKEKLVDDLAFIYHPERLGKPN 505
Db 1079 KQVEEKK---SVKSSSEKVEKESTTTTNDDESEGE-----LSEPTTTTKKD 1123
QY 506 SQIETEDEVRIAQLADKYTSDGYFDEHDIISDEGDAYVPHMGSHWIGKDSLDEK 565
Db 1124 DSKLEPTDEKLISSVPTTAAVEQ-----SRDEFKELE 1156
QY 566 KYAAQAYTEKGLPPSPADVYKANPTGDSAAA-IYNRVKGKRIPLVRLPYVETVEY 624
Db 1157 MDTKESSEK-----KSSPTTTAAASESVKPIDEEKSPPTT-----TTT 1198
QY 625 KGNGLIIP-HKDIHYNIKPFAMFDDHT-----YKAPNGYTLDFATIKYVEHPD 673
Db 1199 TNTTYEPHKKKESKN---DDTTTNTTTTTSKAKSPNSPTRS-----D 1242
QY 674 E--RPHSNDGWNASHVLGKKDHSDD---PNKNFKADEEVEETPA-----EPEVPOV 722
Db 1243 EYVEPHQ-----DASQDEINKRKLEDITSTPSKRLKPDSTPSSATTASTPSEQESP-L 1296
QY 723 ETEKYEAQLKEAEVLLAKYTDSSIKANATETLAG 756
Db 1297 KKENPVGETLSPETIKDSSSSSSSSSSSSSTNTG 1330

RESULT 15
015754 PRELIMINARY: PRT: 565 AA.
AC 015754: 015700;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DR 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
GN WIMA.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Loomis W.F., Iranfar N.;
RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF020407; AAC16025.1; -
FT NON_TER 565
SQ SEQUENCE 565 AA: 62729 MW: B059B6E08201E1CF CRC64;
```

```
Query Match 4.3%; Score 178; DB 5; Length 565;
Best Local Similarity 20.2%; Pred. No. 0.054;
Matches 85; Conservative 60; Mismatches 158; Indels 118; Gaps 16;

QY 24 KQATQKTEML-TPDEYSKREGINAEOIVIKITDQ-----GVVTSBGDIHYHYNCKVP-- 74
Db 118 EQPTQFFSNLIQIPNEY-----TELQSDVIQHOQVYIEHONSITIEQHQRHQFSSDSSNI 172
QY 75 --YDAIISEELMKDPNYKDKEDIVNEYKGGYVVKVDGKYVYVYLKDAHADNVRTKERI 132
Db 173 GIYDSLIEHPILVSD-----NVYNSLSDSITVGKDDDETPNNNNNNIDNIDNIDNI 224
QY 133 NRQKOEHSOHRGCGTPRNDGAVAL-----ARSGRTYTTDGYIFNASDIIEDTGDAYI 185
Db 225 DDNNNTNNNNNSGNQENENLSTSSSEHTPPAQSEPIYVTTTSGSNNNNNNNNI----- 276
QY 186 VPHGDHYIIPKNELSASELAAEAFLSGRGLNS-----NSRTYRQNSDNTSRTNWPVS 241
Db 277 -----INNNE-----NISRDDNSNSNNNNNNNNNNNNNNNNNNNNNN 310
QY 242 SNPGTTNTSNNSTNSQASQSDNDIDSLKQLYKLPISQHVESDGLVF----- 291
Db 311 NNNNNNNNNNNNNNNNNNNNNNGEKEKEISFFS-----SDQVNNSEVLFNSTNSDTNS 363
QY 292 -----DPAQITRTARGVAVPHGDHYHFIPIYQMSLEIRIARIIPLRYRSHHWVDSRP 346
Db 364 SANSIETTVSVTS---AIAAEKDASSF--FNNISASOE-----LQP 399
QY 347 E-QSPQTPPEPSPGPAPAPNLKIDNSSSLVSQLVKVGEGYVEEKGISRYVFAKDLPS 405
Db 400 QPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 457
QY 406 E 406
Db 458 Q 458
```

Search completed: November 13, 2002, 04:10:13  
Job time : 93 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using sw model

Run on: November 13, 2002, 04:08:41 ; Search time 27 Seconds  
(Without alignments)  
867.431 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKEKNRVS.....KLALLKGNPSSYSKEKIN 796

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	US-08-961-083-56	Sequence 56, Appl
2	2649.5	63.6	763	US-08-961-083-66	Sequence 66, Appl
3	1187.5	28.5	447	US-08-961-083-182	Sequence 182, App
4	163	3.9	3696	US-09-134-001C-5080	Sequence 5080, Ap
5	159	3.8	10182	US-09-134-001C-3159	Sequence 3159, Ap
6	147	3.5	812	US-09-513-783A-6	Sequence 6, Appl
7	146	3.5	783	5231168-2	Patent No. 5231168
8	145.5	3.5	1073	US-09-541-782-6	Sequence 6, Appl
9	145.5	3.5	1073	US-09-723-820-6	Sequence 6, Appl
10	144	3.5	1040	US-08-961-083-118	Sequence 118, App
11	141.5	3.4	2314	US-09-816-703A-2	Sequence 2, Appl
12	141	3.4	2431	US-07-920-281C-2	Sequence 2, Appl
13	141	3.4	2431	US-08-466-277-2	Sequence 2, Appl
14	138	3.3	1503	US-08-976-255-14	Sequence 14, Appl
15	137.5	3.3	1610	US-08-513-783A-22	Sequence 22, Appl
16	136.5	3.3	1183	US-08-447-031A-2	Sequence 2, Appl
17	136	3.3	812	US-09-513-783A-4	Sequence 4, Appl
18	136	3.3	1056	US-09-513-783A-32	Sequence 32, Appl
19	134	3.2	2308	US-08-015-973-1	Sequence 1, Appl
20	134	3.2	2308	US-08-448-164-1	Sequence 1, Appl
21	134	3.2	2308	US-08-081-929-2	Sequence 2, Appl
22	133.5	3.2	1664	US-09-599-652-2	Sequence 2, Appl
23	133.5	3.2	1664	US-08-642-846-2	Sequence 2, Appl
24	133.5	3.2	1664	US-09-264-604-2	Sequence 2, Appl
25	132.5	3.2	1507	5268270-2	Patent No. 5268270
26	132	3.2	1164	US-09-457-708-2	Sequence 2, Appl
27	132	3.2	2032	US-09-071-035-458	Sequence 458, App

28	132	3.2	2032	4	US-09-071-035-462	Sequence 462, App
29	132	3.2	2032	4	US-09-071-035-466	Sequence 466, App
30	131	3.1	1093	5	PCT-US93-03077-1	Sequence 1, Appl
31	130	3.1	2115	4	US-09-324-867-5	Sequence 5, Appl
32	129	3.1	2860	2	US-08-826-267-2	Sequence 2, Appl
33	128.5	3.1	916	1	US-08-188-228-48	Sequence 48, Appl
34	128.5	3.1	916	1	US-08-332-643-42	Sequence 42, Appl
35	128.5	3.1	916	1	US-08-332-638-48	Sequence 48, Appl
36	128.5	3.1	1848	4	US-08-296-791-6	Sequence 6, Appl
37	128.5	3.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
38	128	3.1	783	4	US-09-513-783A-176	Sequence 176, App
39	127.5	3.1	2133	2	US-08-670-707A-37	Sequence 37, Appl
40	127.5	3.1	2133	4	US-09-037-601-37	Sequence 37, Appl
41	127.5	3.1	2133	4	US-09-315-179-37	Sequence 37, Appl
42	126.5	3.0	1181	2	US-08-488-940-2	Sequence 2, Appl
43	126	3.0	688	3	US-09-141-047-8	Sequence 8, Appl
44	125.5	3.0	2133	4	US-09-523-656-30	Sequence 30, Appl
45	125	3.0	868	1	US-07-864-004B-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-08-961-083-56
: Sequence 56, Application US/08961083
: Patent No. 6159469
:
: GENERAL INFORMATION:
: APPLICANT: Choi et. al.
: TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
: NUMBER OF SEQUENCES: 452
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,083
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 56:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 796 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-961-083-56

Query Match 100.0%; Score 4163; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 1.7e-314;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYELGLYQARTVKEKNRVSIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGYVT 60
Db 1 SYELGLYQARTVKEKNRVSIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGYVT 60
```



[illegible]

```

Db 1544 IANVL-----PATAVKSKAKKDIDQKLAQIQNIQTHQTATTEE 158

```

```

QY 326 RIARIIPLYRSNH-----WVPSRPPSPQPTP--- 355
: : : : :
Db 1583 KEAAIOLANOKSEARFATAIONEHNSNGVAAQAKSNGIHEIELVMPDAHKKSDAQSIDNKY 1642
QY 356 -EPSPPOAPNLIKIDSNSSLSVQL--VRKVGEGYFEEKGISRYFAADLPSEYKNE 412
: : : : :
Db 1643 NQDSNTINTTPATDEKOKALDKLTIARDAGYNKVDQAOTNOQVSADATEALDITITNIQ 1702
QY 413 SKLSQESVSHLTAKKE-----NVAPRDOE-----FVDAKYNLL----- 447
: : : : :
Db 1703 ANVAKPSPARVELDSKFEDLKRQINATPNATEBEKODATQRLNGKDEKYNKLNIDRRDN 1762
QY 448 -TEAHK-----ALFXNKGKNSDFQALDKLERL-----NDESTNKEKLVDDL 489
: : : : :
Db 1763 EYEOHKNIGLOELETIHAPTRKSD--ALQELQTKFISQTELINNNKDKATNEEK---DEA 1817
QY 490 AFLAPITHEERL-----GKPNQIEYTEDE--VRINQL-----ADKYTTSQGYTFDEH 535
: : : : :
Db 1818 KRLETSKKNKTTTNNIQAOTNNQVNAKONGMNEIATITIPATTIKTDAKTALDKKAEQV 1877
QY 536 DIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQYTKGILPPSPADYKANPTGDS 595
: : : : :
Db 1878 TITNGNMDA-----TDEKAPARKLVEKAKI-----EAKSNTITNSD 1913
QY 596 AAATYRVKGEKRIPLVRLPYWEHTVEVKNGLIIPKDHYNHKEFAMFDDHTYKAP-- 653
: : : : :
Db 1914 TE---REVNGAKTNGLEKINNIOPTQTKNAKQEIINDKAQEQLIQINMTPDATEEKQOE 1970
QY 654 -----NGYTLEDL-----PATIKYVEHDERP---HSDMGKGNNS 686
: : : : :
Db 1971 ATNRVNAQLAQAIONINNAHSTOEVNESKTSIATIKSVOPNKKPRTAINSLTQEBANNO 2030
QY 687 EHYLKGKHSDEPNKMF-----KADE--EPYEETPAPEVYPOVTEKEVEA-----Q 730
: : : : :
Db 2031 KTLIGMDNAGTADDEKFAAKQVLTQKLEQIKHESTONQVNVNAQAQITAKKLINANA 2090
QY 731 LKEAEVL-----LAKYTDSSLAN--AT-----ETLAGIRNNL---TLQIND 767
: : : : :
Db 2091 HKRODINILITLAEKSKDIRANQDATTETEEKTAIOSIDTFLAQRNNINCAANTALVD 2150
QY 768 NNSIMAEKLLALLKGSNPSYSKEKI 795
: : : : :
Db 2151 EN-LEDGKQRLQRIVLSTQTKTOAKADI 2177

RESULT 5
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

Query Match 3.8%; Score 159; DB 4; Length 10182;

Best Local Similarity 18.8%; Pred. No. 0.044;

Matches 189; Conservative 126; Mismatches 368; Indels 320; Gaps 43;

```

QY 13 KENNRVSYIDGQAOIKQNTENLTPDEVSKREGIN-----AEQIVIKITDQGYVTSH 62
: : : : :
: : : : :

```

```

Db 5419 OKSGEESLVNSNTRSEVE-----EHLNEAKSLNNAMKQLRDKVAEKTNYK--OSSDYINDS 5473
QY 63 GDHYHYNGKV--PYDAIISE-----ELLMK-----DPNYKLIKDEDIVNEVGQGYI 107
: : : : :
Db 5474 TEHQKGYDQALQEAVENTIEIGNPILNKSEIEOKLQQLTDAQNALQSGHLEAKNNAIT 5533
QY 108 KYDGKYYVYLKDAH--ADNVRTKEEINROKQESHQREGTPRND-----GAVALARIS 159
: : : : :
Db 5534 GIN--KTLFALNDAQRQKALENVQAQOTIPAVNQQLTLDBREINTAMQALRDKYGGQGNVHQ 5591
QY 160 QGRYTTDDGYIFRNASDIETDGTATVPHGDHYHIFPKNEL--SASELAAPAFISGRGN 217
: : : : :
Db 5592 QSNYFEMEDQPKHNDVNSVQAQOTIIDLQD--PIINKHEIQALQINMTQTALSGENK 5649
QY 218 L-----SNSRTYRORSDNTSRTNWPVSNSPQTTNTSNNSTNSQASQSDNDISLTK 272
: : : : :
Db 5650 LHTDQSTNRQLEGLSSLTWTQAQINAEKDLVNAQKTTDVA-----QKLAKEINSAMS 5703
QY 273 QLYKLPLSQRHVESDGLVDPQAQITSRTARGVAVPHGDHYHIFYSQMSLEERLARIIP 332
: : : : :
Db 5704 NL-----RDGI-----QNKEDIKRSSAYINADPTKYTAYDQALQNMENIINATP 5747
QY 333 LKRSNHWVPDSRPEQSPQPTPEPSPQAPANLKIDSNSSLSVQLVKVGEQYFEEK 392
: : : : :
Db 5748 -----NVELNKATIEQALSRYQQAQO--DLD 5771
QY 393 GISRYVEAKDLPESEYKNLES-----KLSQESVSHLTAK 428
: : : : :
Db 5772 GYQQLANAKQAQATQYVNGLSLNDQKRELNLINSANTRTYQOEELNKATELNHAMEL 5831
QY 429 KENVAPRD-----QEPYDKAYNLLTEAHKALFXNKGKNSDFQALDKLER 473
: : : : :
Db 5832 RNSVQVWDQKQSSNYVNEDEQPHQHNVDYAVN--EAQATINNNAQPVLDKLAIERLTQT 5888
QY 474 LN---DESTNKEKLVDDLAFAPITHEERLGRKPSQIEY-----TDEVR-IAQ--- 519
: : : : :
Db 5889 VNTTKALHGAQKLTQDQQAETGIRGLTSLNEPQNAEAVATATTEDEVARNIRQEAAT 5948
QY 520 -----LADKYTT--SDGYTFDEHD-----TISDGDYAVTPHMGH 552
: : : : :
Db 5949 TIDTAMLGKRSIKKQNDPNKNSKYINEDHDQQAQAVDNAVNNAAQOVIDETQATLS----- 6003
QY 553 SHWIGKDSLDEKVAQAQYTKGILPPSPADYKANPTGDSA-----AAIYNRVKGEKR 608
: : : : :
Db 6004 -----SDTINOLANAVTOAKSNL-----HGDFTKIQHDKDSAKQTIQALQNLNSAQKH 6050
QY 609 -----TPIVRLPYMVEH-----TVEVKNGLI--IPKDHYN-- 638
: : : : :
Db 6051 MEDSLIDNESTRTOVOHDLTEAQLDGLMGALKESTIKDYTNIVSGNYTINAEPSKKQAYD 6110
QY 639 -----NIKFAMFDDHTY--KAPN--GYTLEDL--- 661
: : : : :
Db 6111 AAVQNAQNIINGTNOPTIKGNVTTATQTVKMTKQALDQDHLLEKKNNAQNTIRNLSL 6170
QY 662 -----PATIKYVEHDERPHSNDGWSAEHVLGKDH-----SEDP 699
: : : : :
Db 6171 NNAQDAEKNLVNSASTLBOVQOQNLQTAQQLDNPANGELRQSI--AKKQDYKADSKYLINEP 6229
QY 700 KKNFKADE-----EPYEETPAPEVYPOVTEKEVEQAQAEVLL--AKYTDSSLANATE 752
: : : : :
Db 6230 QIKQNYDQAVQREVTIINETQNPBELKANIDQATOSVQNAEBALGAELKALNODKOTSSTE 6289
QY 753 TLAGLRNNLTLOIMNNSIMAEKLLALLKGSNPSYSKEKI 795
: : : : :
Db 6290 -LDGL-TDLT-----DAQREKLREQIINTSNSRBDIKOKI 6321

```

RESULT 6  
US-09-513-783A-6  
; Sequence 6, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.

```

1  APPLICANT: Kapur, Ravi
2  TITLE OF INVENTION: A System for Cell Based Screening
3  FILE REFERENCE: 97-022-L1
4  CURRENT APPLICATION NUMBER: US/09/513,783A
5  CURRENT FILING DATE: 2000-02-25
6  NUMBER OF SEQ. ID NOS: 180
7  SOFTWARE: PatentIn Ver. 2.0
8  SEQ ID NO 6
9  LENGTH: 812
10 TYPE: PRT
11 ORGANISM: Artificial Sequence
12 FEATURE:
13 OTHER INFORMATION: Description of Artificial Sequence
14 OTHER INFORMATION: EYFP-DEAD-MAPKDM construct
15 US-09-513-783A-6

```

Query Match	3.5%	Score 147	DB 4	Length 812
Best Local Similarity	19.8%	Pred. No. 0.0068		
Matches 175; Conservative	103;	Mismatches 331;	Indels 274;	Gaps 39

0Y	53	ITDQY-----	-VTSGBDHHYNGVAPYDAI	ISELLMKCD--	NKRLDE-----	-DIV	98							
Db	62	VTTFGYGLQCFARY	PDHMKQHDFFKSA	MPGVOERTIPFKD	GNKTTAREVYFEC	DVLY	121							
0Y	99	NEVK-GGVYIKVDG-----	-KIYYLKA	HAHADNVKTEEL	NROKQHS----	-Q	141							
Db	122	NRIELKIDGFEKD	ENILGHKLE	VYNYSNHYIMADK	OKNGIKVNFKIRHNI	EDGSVOLAD	181							
0Y	142	HREGTPNOC	AVALLASQGGYTTD	DOCTIN	ASDIIIEDTC-----	-AYIVPAG	189							
Db	182	HYQNFPTG	DBSPVLL-----	-PDNHYLSY	QALSXKDPNEKRDH	VLLFYTAA	233							
0Y	190	DHYIYPAN	ELASSEL-----	-AAAA-----			210							
Db	234	MDELXKPR	DEADSDLS	LYDALTPPE	ELIGELIKRODMAL	LEAPPDVIG	VEKTEETI	293						
0Y	211	-FLSGRGLNS	RTYRRONSD	NTSRTN	MVPSVSNPGTNTNTS	NSNSQASQ	SDIDS	269						
Db	294	PULDGEK	TGSESKKKPC	LD-TSQV	EGIPS-SKP-TLLANG	DGHGEGN	TAGSPD--	347						
0Y	270	LLKQYKLP--	-LSORV	ESGLVFDP	QAQITTSR	ARKVANPHOD	DHFIPI	QSGMSLEBR	326					
Db	348	FLEEVDY	PDYQSSON	MPEDASFQ	QOQVLD	PDQ--	-AEFENEH-----	-RDG	393					
0Y	327	IARIPL	KRSNMHW	PPSRPEOP	POPPTPPSPG	OPAPYLKIDS	NSLSVQ	LVRKVGEG	386					
Db	334	LADLLF	VSSTG	TNSAP	TERDNPS-----	-EDSGMLPC	DSF--ASTAY	VSQMSVGA	GN	445				
0Y	387	YVEEKG	ISRYVFAK	DLPS	ETVKNLE-----	-SKLSKQ	ESVSHTLTAK	KENVAR	PDQEFYD	441				
Db	446	SPCS	SCVVS-----	-PEVTI	ETLQAT	ELTSLAA	EVESYK	EQDLPAK-----		484				
0Y	442	KAYNL	LEPAHKA	LFXNKG	GRNSD	FALDKLLER	LAND-----	-ESTKKE	LVDLLA	FLAPITH	497			
Db	485	-----									519			
0Y	498	PERLGP	NSQLE	TEDE	RYLAQ	ADKTTSD	GVYFDE	HOIDS	EGDAY-----		545			
Db	520	-----									568			
0Y	546	-----									588			
Db	569	SSTKDV	APRME	EELVP	OND	TTSPRE	KTETI	PLIKMD	LAPRED	VLLTKETELAP----	AKGM	624		
0Y	589	ANP	GSAA	ALY	NRVKG	ERIP	VLRLPY	VEHYEV	YKNGMLIIP	HKDHYH	INIK	FAFDDH	648	
Db	625	VLSL	EEIALK	KNVRS	AE-IPVAQ-	ETVVS	ETEV	LATE	VLPL-----			666		
0Y	649	TYKAN	NGTLE	DLATIK	YIYYEH-	-DER	HSND	GMQNS-ASE	HYLKKD--	HSD	PKNKK	704		
Db	667	-----										716		
0Y	705	ADEBE	VEETPA	-----	-PEV	QVTE	KVEV	QKAE	VALILAK	VTDSS	LKAN	ATE	FLAG	756

[illegible]

Query Match 3.5%; Score 146; DB 6; Length 783;  
Best Local Similarity 20.2%; Pred. No. 0.0077;  
Matches 141; Conservative 106; Mismatches 278; Indels 174; Gaps 35;

[illegible]

```

RESULT 8
US-09-541-782-6
; Sequence 6, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-541-782-6

```

```

Query Match          3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.014;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

```

```

QY 23 GQAOTKTEMLTPDEVSKREGINAE-----QIVIKI--TDQGYVTSBGDHYHYNN 70
DB 147 GGTGCTKTYTMSGD-LSDSGDIISGAGLIPRALYQLFSSLDNSQGEYAVK-CSYEELYN 204
QY 71 GAVPYDAIISSELMKDPYKDKDED-----IVNEVKGGYVKKVGGKYVYLKDAAHADN 125
DB 205 EEI-RDLIVSEEL--RKPARVFEEDTSRGNVYVTGIEEY--IKNAGDGLRLREGSHRQ 260
QY 126 VTKKEINRQKQEH-----OHREGTPRNDGAVALARSOGRTTDDGYIFNASDIIDET 180
DB 261 VAATKCNDSLSSHSIFTLTKRKVSSGMDTETNSLTINN-----NSDDLRL-- 307
QY 181 GDAIYVPHGDHYHYIPKNELSASELAAPLSSGRLNSNRT----- 223
DB 308 -----ASKLHWVDLAGSENI--GRSGAENKRARETGMINOSLITLGRVY 349
QY 224 -----YR-----RQNSDNTSRTMWVPSVSNPGTTNTNTSNNSTNSQASOS 264
DB 350 NALVEKAHHIPYRESKLTRLLDOSLGKTKTSMIVTVS--STNTNLEETISTLEYAARA 406
QY 265 NIDSLKQLYKLPISORHVESDGLVFPDAOI-----TSRTARGVAVPHGDHYHFIPYSO 319
DB 407 K---SIRKRPQNNQVLFKRVLLKDLVLDIERLKNLAINATRKKNGYVLAESTYKELMDRVQ 463
QY 320 MSEL-----EERTARIITPLKRYSNHWVPSRPEQSPQPTPEPSPGPOPAPNLKI----- 369
DB 464 NNDLLCOEQBARLEVLDLNVKS-----SRDLOQYVSKSNQEHKKKVEALQLOLVNSSTE 517
QY 370 -----DSNSSLVQLVRAKYGEGIVFEKGISRYVAKDLPSETVKNLESKLSKOESVSHT 424
DB 518 LBSVSENENKLNELVLEIEKRRKYETNEAKITVATDLSQYRRESKEYIASIYEXELDT 577
QY 425 LPAKKENVAPROEFYDKAYNLL-----TEAH 451
DB 578 EKNKNEN-----ENFWNLKFNLLITMLRSFNGSEFTDETNGYFTLLDNFNASMEELLNTHSN 633
QY 452 KALFXNKNNSDFOALDKLLER-----LNDESTNKEKLVDDLLAFAPIT 496
DB 634 QLLISMTKLTHERQSLDALQASRSCAVPNSSLDLIVSELDSKNSLLDALEHSLODS 693
QY 497 -HPERIGKNS-----QIYTTDE-----VR--IAOL 520
DB 694 MSSOKIGNISELLELQDKMESYRQLOVELRSLYLNLOHTHEESQKELMYGVRNIDAL 753
QY 521 AKKYYTS--DGYIFEDHDIISDEGDAYYVPHMGSHWIKKDSLSDKEKAAQAYTEKEGI 578
DB 754 VKTCTTSLNDADIT-LSDYISDQSKFEESKODLIANIGKIVSNFLQECNESLYTK----- 808

```

```

QY 579 LPPSPDADYKANPTGDSAAAI--YNRKGEKRIPIVR-----LPYWEHTVEVK 625
DB 809 -----ADILHSHLNDNTSNIRKANEMNNRSEEFLERNAASQAEIYGANKERIQKTE-- 860
QY 626 NGNLIIPKHDH-YHNIKRFMFDDHTYKAPNG-----YTLIEDLFAITKIYVEHPDERP 676
DB 861 NSQLDSSKSKAHSNSRSMYDHCALAESQKQVNLLEVOTLDRLLOKVK-----E 911
QY 677 HSDNGKNSSEHYLGCKKHSDPENKNFKADEPVEETPAPEVPQVEPEKVEAOLKEAV 736
DB 912 HSDNTEKKEHQQLDLESLVGNNDVL-----IDSIKYPHTLO-- 950
QY 737 LIAKYVDSLK-----ANATETLAGLRN--NLTLQIMDNNSIAEA-----EKLA 780
DB 951 ---KITDHYLKGTSLANTNTNELLGIGDESLCNLETTIEDTSLVLYLETTGDTPRKRELPA 1007
QY 781 LKGSNPPSSVSKKIN 796
DB 1008 TFSWTRDSSLIKETTN 1023

```

```

RESULT 9
US-09-723-820-6
; Sequence 6, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-723-820-6

```

```

Query Match          3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.014;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

```

```

QY 23 GQAOTKTEMLTPDEVSKREGINAE-----QIVIKI--TDQGYVTSBGDHYHYNN 70
DB 147 GGTGCTKTYTMSGD-LSDSGDIISGAGLIPRALYQLFSSLDNSQGEYAVK-CSYEELYN 204
QY 71 GAVPYDAIISSELMKDPYKDKDED-----IVNEVKGGYVKKVGGKYVYLKDAAHADN 125
DB 205 EEI-RDLIVSEEL--RKPARVFEEDTSRGNVYVTGIEEY--IKNAGDGLRLREGSHRQ 260
QY 126 VTKKEINRQKQEH-----OHREGTPRNDGAVALARSOGRTTDDGYIFNASDIIDET 180
DB 261 VAATKCNDSLSSHSIFTLTKRKVSSGMDTETNSLTINN-----NSDDLRL-- 307
QY 181 GDAIYVPHGDHYHYIPKNELSASELAAPLSSGRLNSNRT----- 223
DB 308 -----ASKLHWVDLAGSENI--GRSGAENKRARETGMINOSLITLGRVY 349
QY 224 -----YR-----RQNSDNTSRTMWVPSVSNPGTTNTNTSNNSTNSQASOS 264
DB 350 NALVEKAHHIPYRESKLTRLLDOSLGKTKTSMIVTVS--STNTNLEETISTLEYAARA 406
QY 265 NIDSLKQLYKLPISORHVESDGLVFPDAOI-----TSRTARGVAVPHGDHYHFIPYSO 319
DB 407 K---SIRKRPQNNQVLFKRVLLKDLVLDIERLKNLAINATRKKNGYVLAESTYKELMDRVQ 463
QY 320 MSEL-----EERTARIITPLKRYSNHWVPSRPEQSPQPTPEPSPGPOPAPNLKI----- 369

```



```

Db 464 NKDLCOBARKLEVLADLNVS-----SRQLOVSKSNQEHKKEVBALOLOLVNSSTE 517
QY 370 -----DSNSLVSQLVRRYGEYVEEKGISRYVPAKDLPESTVKNLESKSKOSVSHT 424
Db 518 LESVSESEKLNELVEIERKKYETNEAKITTYATDLISOYRSKEYIASLYEKLRT 577
QY 425 LTAKENVAPRDOEFYKAYNL-----TEAH 451
Db 578 ERNNKEN-----ENFMNKKFNLLTMRSPHSGFTDETNGYFTLLDNFAMSEMLNHSN 633
QY 452 KALFYKNGRNSPQALDKLER-----LNDESTNKEKLVDDLAFAPIT 496
Db 634 QLLISMRTITEHFOSIDBALQASRSCAVPNSSLDIYSELKSKNSLIDLAEHSLQDIS 693
QY 497 -HPELGRPN-----OLEYFDE-----VR--IAQL 520
Db 694 MSSOKLNGISSELELODKMEKESYROLVQELRSLYNLOHTHEESOKELMTGVRNDIDAL 753
QY 521 ADKYTTS--DGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDERKVAQAAYTKERGI 578
Db 754 YKCTGTSINDADI--LSDYISDQKSKFESKQODLIANIGKIYSNFLQDQNESLYTK----- 808
QY 579 LPSPDADYKANPTGDSAAI--YVRKGEKRIPLV-----LPYVHEHTVEVK 625
Db 809 -----ADLHSHLNDTNSNIRKANEIMNNRSEEFIRNAASQAEIVGANKERIOKTV-- 860
QY 626 NGNLIIIPKDH-YHNIRKAMFEDHTYKAPNG-----YTLFEDFAITIKYVEHPDERP 676
Db 861 NQSOLDISKRAIHSNSRSMIDHCIALAESOKQVNLQVOTDRLOKQ-----E 911
QY 677 HSNDSWGNASEHVLGKSHSEDPNKNFKADEEPVEETPAEPPOVETEKVEAQLKEAV 736
Db 912 HSEDNTKKEHQOOLDLESIVGNNDL-----IDSKIPHELO----- 950
QY 737 LLAQYDSSLK-----ANATETLAGLRN--NLTLQIMDNNSIAEA-----EKLLA 780
Db 951 --KITDVLKGTGTSIANHVTLELLGDESLCNLETTIEDTSLVLETTGDTSPKRELP 1007
QY 781 LKGSNPSSVSKEKIN 796
Db 1008 TPSPWTRDSSLIKETTN 1023

RESULT 10
US-08-961-083-118
; Sequence 118, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders

```

```

; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1040 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-118

Query Match 3.5%; Score 144; DB 4; Length 1040;
Best Local Similarity 19.2%; Pred. No. 0.017;
Matches 154; Conservative 111; Mismatches 284; Indels 252; Gaps 38;

QY 107 IKVDGKYV-YLKDAAHADNVTKKEINRQKQESHQAREGTPRND-----G 152
Db 76 LKIEGYIYGIK-TKQDNTELSRTVDGKXSAQSDQPNSTKTSDDVHASDLEWNOGQ 134
QY 133 AVALARSQGRITTDGTYFNASDIITEDGDAYIVPHGDHYIIPKNELSASELAAEAPL 212
Db 135 KYSL--OGESAGDG-----LSEKSTIADNLSSNDSFA 166
QY 213 SGRGNLSRTRYRQNSDNTSRTNVPSVSNPGTNTNNSNNSNQSASQNDISLKL 272
Db 167 S-----QVEQNPDKHGESVRYPTVPEGQNVYSATTVOSAEVEVLATIND--R 211
QY 273 QLYKLPLSQRHVESDG-----LVF-----DPAQITS 298
Db 212 PEYKLPLETKGTQGPBGHGEAAVREDLPYTKPLETKGTQGPBGHGEAAVREEPAYTER 271
QY 299 RTARGVAVPHGDHYIFITYSQMSELEERLARIPLRYSNHWPVPSRQPSQPTPPS 358
Db 272 LATKGTQGPBGH-----GATVREETLEYTEPAVATKGT-----QEBEHGEKRYVEEL 319
QY 359 PGPQPA-----PNL-----KIDSNSLVSQVLR--KVGEGYVEE 391
Db 320 PALEYTTNRNRTIOMIPTTEIOPITLKNRKLIERQOAGITRIQYEDYIVNGVYET 379
QY 392 KGISRYVPAKDLPESTVKNLESKSKOSVSHTLTAKENVAPRDOEFYKAYNL--TE 449
Db 380 KEVSRTEVA--PVNEVYKVGTLVAKPVELTNLTKEYN-----KKSITVSYNLIDTTS 431
QY 450 AHKALFYKNGRNSPQALDKLERLNDESTNKEKLVDDLAFAPIT--TH-PELGRPN 505
Db 432 AYVS-----AKQYVFHG-DKLVEVDIENPAKEQVIG-LDYTPPYTVKTHLTLYMGENN 484
QY 506 SQIETDEVRIAOLADKYTTSIDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLQKE 565
Db 485 E--ENTETSTQDFOL--EKKIETIDISVELYGENRY-----RRYL--SLSEAP 530
QY 566 KYAQAAYTKERGIPLPSPDADYKANPTGDSAAIYNRYKGEKRIPLVRLPYVEHT----- 621
Db 531 TDTAKYFVKV-----SDRFKEMLPVKS-----ITENTDGT 563
QY 622 -VEVKNGLIIPKCHYINIKFAFDHTY-----KAPNGYT--LEDLFAITK-----Y 667
Db 564 KYTAVAVDQLEVEGTGKY-----DGYTFYAKSKAQPGYYSFKOLVYAMQNSLGV 616
QY 668 YVEHDERPHNSDQGNASEHYLGKKDS-----EDPKNRKADE--EPVEET 713
Db 617 YTLASDM-----TADVSLGDKQTSYLTGATGSLISGDSGKTSYAIYDLKAPLPFD 667
QY 714 PAEPVPOVETEKVEAQLKEAEVLLAKYDSSL-----KANATETLAGLRNLLQI 765
Db 668 LNCATVRLDLIKTVSADSKENVAALAKAANSANINNAVEGIGSAGKAVGIVASATNTV 727
QY 766 MDNNSIMAEAEKLLALKGSN 786
Db 728 IENSSFTG--KLIANHDSN 745

```

```
RESULT 11
US-09-816-703A-2
; Sequence 2, Application US/09816703A
; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 AGY
; CURRENT APPLICATION NUMBER: US/09/816,703A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Krueger, N.X. and Saito, H.: A human transmembrane protein-tyrosi
; OTHER INFORMATION: ne-phosphatase, PTP zeta, is expressed in brain and has an N-term
; OTHER INFORMATION: inal receptor domain homologous to carbonic anhydrases; Proc. Nat
; OTHER INFORMATION: l. Acad. Sci. USA 89 (16), 7417-7421 (1992)
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Levy, J.B., et al.; The cloning of a receptor-type protein tyrosi
; OTHER INFORMATION: ne phosphatase expressed in the central nervous system; J. Biol.
; OTHER INFORMATION: Chem. 268 (14), 10573-10581 (1993)
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION: By similarity
; NAME/KEY: DOMAIN
; LOCATION: (25)..(1635)
; OTHER INFORMATION: Extracellular (potential)
; NAME/KEY: misc.feature
; LOCATION: (25)..(2314)
; OTHER INFORMATION: Mature chain; protein-tyrosine phosphatase zeta
; NAME/KEY: DOMAIN
; LOCATION: (34)..(302)
; OTHER INFORMATION: Carbonic-anhydrase like
; NAME/KEY: SITE
; LOCATION: (105)..(105)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (134)..(134)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (223)..(223)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (232)..(232)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: DOMAIN
; LOCATION: (312)..(406)
; OTHER INFORMATION: Fibronectin Type-III
; NAME/KEY: SITE
; LOCATION: (324)..(324)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (381)..(381)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (497)..(497)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (501)..(501)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
```

```
; LOCATION: (552)..(552)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (587)..(587)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (602)..(602)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (629)..(629)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (637)..(637)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (677)..(677)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: VARSPLIC
; LOCATION: (755)..(1614)
; OTHER INFORMATION: Splicing variant; missing (in short isoform)
; NAME/KEY: BINDING
; LOCATION: (997)..(997)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1017)..(1017)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1050)..(1050)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1082)..(1082)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1122)..(1122)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1456)..(1456)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (1548)..(1548)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: BINDING
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1561)..(1561)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1617)..(1617)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: TRANSMEM
; LOCATION: (1636)..(1661)
; OTHER INFORMATION: Transmembrane region; potential
; NAME/KEY: DOMAIN
; LOCATION: (1662)..(2314)
; OTHER INFORMATION: Cytoplasmic (potential)
; NAME/KEY: CONFLICT
; LOCATION: (1722)..(1728)
; OTHER INFORMATION: Missing (in ref. 2)
; NAME/KEY: DOMAIN
; LOCATION: (1744)..(1997)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: ACT_SITE
; LOCATION: (1932)..(1932)
; OTHER INFORMATION: Active site; by similarity
; NAME/KEY: DOMAIN
; LOCATION: (1998)..(2314)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: misc.feature
; LOCATION: (2222)..(2222)
; OTHER INFORMATION: Ancestral active site
US-09-816-703A-2
```

Query Match	Similarity	3 4%: Score 141.5; DB 4; Length 2314;
Best Local Match	19.2%: Pred. No. 0.095;	
Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 433		
QY 51	IKITQGVYVTHSGDHVHYHYNGKVPYDAIISEELMKKDPYKJLKDIEDIVNEVGKGYIKVD	110
Db 954	VGVTYQGSILFSPSHII-----PIPKSLITTPASLLQPHALISG-----D	993
QY 111	GKRYVYLDDAAHADNVRKKEELINROKOEHSOHEGTPRNDGAVAL-----ARSGRYT	164
Db 994	GEM-----SGASDS-----EFLLPDTGLFTLNLISPVSAVEFTYT	1030
QY 165	T-----DDGIFPASIIEDTGATLYVPHGDHY-----IPK-----NELSASELAAAE	209
Db 1031	TSVFEGDDKKALSKSEIIYGNETELQIPSENEVWYSESTVMNMYDNVKKLNASLQETSV	1090
QY 210	AFLSGRG-----NLSRSRYR-RQNSDNTS-----RTNWVP	239
Db 1091	SISSTKGMFPGLATTTKVEDHETISOYPENNFSTQPHHTYSQASGDSLKLKLVLSANSEP	1150
QY 240	SVSNBGT-----NTNTSNNSNTNSQAS-QSNDIDSLKOLYKILPLSORHVES	286
Db 1151	ASSDPASSEMILSPSTQLLEFETSSAFSTREVLQPSFQASDVDTLLKTV--LFA-----VPS	1204
QY 287	DGLVDPDQO-----ITSRARGVAVPHGDHYHPIYSQMSLEERARIIPLRY	335
Db 1205	DPILVETPKYDKISITMLHLIVSNSASENMHSTSVPFVDVSPSTSHMSASLQGLTISY	1264
QY 336	RSNHVVPDSRPEOPSPOTPEPSPPQAPANLKDINSNSLSQLVRKQEGVFPBEKGS	395
Db 1265	ASEKEPEVLKSESSHQVPP-----SLYSNDELFOGTANLEIQAHH--PPKG--	1308
QY 396	RVPFAKDLPS--ETVKNLKESKLKQESVSHLTJAKENYAPR-----DOEEFYDAY	444
Db 1309	RHVETPVLSTIDEPILNTLILNKLIHSDET--LTSKSSYTGAVFAGIPIVASDITVSDIH	1365
QY 445	NL-LTEAH-----KALPFXNG-----	459
Db 1366	SVPIQNGHVAITAASPHRDGSVYTSYTKLLFPFSKATSELSHSAKSDAGIVGGEDGDTDDGC	1425
QY 460	-----RNSDPQALDKLLRLNDESNKKKIYVDLLAFIAPITHPERLQPNQIEX----	510
Db 1426	DDDDRRDSDGSIHFKCMS--CSSYRESQEKVMD-----SDTHENSLMDQNPISISISE	1478
QY 511	-----TEDEVRIQAOLADKRYT-----SDQYFDEHI-----ISDEG	542
Db 1479	NSEEDNRTYVSSDSQGTMDRSPGKSPSANGLSQKHNDGK--EENDIQTGSALLPLSPES	1536
QY 543	DAYVTPPHMGSHHWIG--DSLSDKEKYA--QAAYTKER--GIL-----PPSP	583
Db 1537	KAMAVLTSDDESGSGOGTSLINENETSTDPFSFADJNEKADGIIAAGDSSETTPGPQSP	1596
QY 584	DADV-----KANPTGSAALYINRVGGER--IPVLRPLRMVHTVEYANG	627
Db 1597	TSSVTSENSEVFPVHSEADASNSHESRIGLAEGLSEKRAVPLIVYSALFICIVLWG	1656
QY 628	NL-----IIPKHQVHNKIFAMFDDHT--YKAPNGY	656
Db 1657	ILIIWRKCFQTAHFLELDSTSPRVISTPTPTPIFISDVGALPPIHFKHVAHDLAASGF	1716
QY 657	TLEDLEFATIK-YVE-----HDERPHSN	679
Db 1717	TEE--FETLKEFYQEVOSCTVLDGITADSSNHPDK--HKN	1753
RESULT 12		
US-07-920-281C-2		
; Sequence 2, Application US/07920281C		
; Patent No. 5739026		
; GENERAL INFORMATION:		
; APPLICANT: Garoff, Henrik		
; APPLICANT: Liljestrom, Peter		
; TITLE OF INVENTION: DNA Expression Systems Based on		

1 TITLE OF INVENTION: Alphaviruses  
 2 NUMBER OF SEQUENCES: 27  
 3 CORRESPONDENCE ADDRESS:  
 4 ADDRESSEE: Birch, Stewart, Kolasch & Birch  
 5 STREET: P.O. Box 747  
 6 CITY: Falls Church  
 7 STATE: Virginia  
 8 COUNTRY: USA  
 9 ZIP: 22040-0747  
 10  
 11 COMPUTER READABLE FORM:  
 12 MEDIUM TYPE: Floppy disk  
 13 COMPUTER: IBM PC compatible  
 14 OPERATING SYSTEM: PC-DOS/MS-DOS  
 15 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 16 CURRENT APPLICATION DATA:  
 17 APPLICATION NUMBER: US/07/920,281C  
 18 FILING DATE: 13-AUG-1992  
 19  
 20 CLASSIFICATION: 435  
 21 ATTORNEY/AGENT INFORMATION:  
 22 NAME: Murphy Jr., Gerald M.  
 23 REGISTRATION NUMBER: 28,977  
 24 REFERENCE/DOCKET NUMBER: 828-103P  
 25 TELECOMMUNICATION INFORMATION:  
 26 TELEPHONE: 703-241-1300  
 27 TELEFAX: 703-241-2648  
 28  
 29 TELEX: 248345  
 30  
 31 INFORMATION FOR SEQ ID NO: 2:  
 32 SEQUENCE CHARACTERISTICS:  
 33 LENGTH: 2431 amino acids  
 34 TYPE: amino acid  
 35 TOPOLOGY: linear  
 36  
 37 MOLECULE TYPE: protein  
 38  
 39 US-07-920-281C-2

Query	March	Similarity	3.4%:	Score 141:	DB 1:	Length 2431:
Best	Local	Similarity	19.7%:	Pred. No. 0.11:		
Matches	137:	Conservative	93:	Mismatches	253:	Indels 212: Gaps 34:
QY	56	QGVYVSHGDHYYNG-KVYPDAITSEELIMKDPYKLIKDEIVINEVGGYIKYDKKY	114			
Db	1523	KGYSTTDSLSYSEFGCTFENQAIMAELTIMPLQENEDICIXALGETM-----	1574			
QY	115	YLLKDAAHADNVRTKEELNRKOEHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNAS	174			
Db	1575	-----DNIRSKCPVN-----DSDSTPRT-VPCLCR-----YAMAE	1606			
QY	175	DIIDETG-----DAIYVPHGHHYIKNEISASBELAAAEFLSGRLNSRITYR	225			
Db	1607	RIARLSHQVKSVMVYSSFPPLP-KYHVDGQKCKECKVLLFDETPV---SVSPRXYA	1660			
QY	226	RONSNTSRT-----NMVPSVNGTNTNTNNSNINSQASNDIDSLIKOLYKPL	279			
Db	1661	ASTDHSRSLRGFLDM-----TTDSSASTDMSLSLQSCDDSDIIEPAPIVY	1712			
QY	280	-SQRHVESDGLVFDPQAITSRTARGVAVPHGDHFIPIYSQMSLEERIARIILPRYSN	338			
Db	1713	TADVHPEDAGIADLAADVHPEA-----DHV-----DLENIP-PPRPRA	1752			
QY	339	HWPDSRREQSPQTPPSPQOPAPMLKI-----DSNSLSVQLYKRVGEGYFEEKG	393			
Db	1753	AYLASRAAEREPAPAR-KPTAPAPAFANFKLPITFGDEDEHEVDALASITFGDEDDVLR	1811			
QY	394	ISR---YVEAFKDPETVKNLESKTSKQESVHTLTAKKENAVAPDQEGYKAVYLLTEA	450			
Db	1812	LGRGAYITFSSDTGS---GHILOQSKVRO---HNIQCAQLDVOBEKMPPK--LDTER	1861			
QY	451	HKALE-----XNKGRNSDFQALDKLLERLNDESTNKEKYVDLLAFLAP--IYHP	498			
Db	1862	EKLILLKQMHMPSEANKSRYS-RRVEMMKATVYDRLSGARLYGADVGRIPYAVAYP	1920			
QY	499	ERLQKPNQSIETYEDVIA---QLADKYTTSBGY-IFDEHDIISDEGDAYVYPMHGS	553			
Db	1921	RPVSPVPIEERFSSDVAIACNEIISNYPVAVSYQITDEDAVLDVND-----	1970			



REFERENCE/DOCKET NUMBER: 229/182  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-976-255-14

Query Match 3.3%; Score 138; DB 4; Length 1503;  
Best Local Similarity 19.4%; Pred. No. 0.09; Mismatches 276; Indels 228; Gaps 38;  
Matches 150; Conservative 118;

QY 34 TPDEYSKREGINABOIVIKITDQGYVTSBGD-----HHYNG-KVPYDAIISELL 84  
DB 582 TGPESLQTLARVSELESSTDEDFQSSDPKDSLPGDLHVSGSPSPNNIF----- 636  
QY 85 MKDPYKIKLKDIDYNEKGGVYKVDGKYVYLDAHADNVKTEINROKOEHSQARE 144  
DB 637 ---NDVKSSEDLPSSHOKIFDIMEING-VQADFPRATLSSLD-----NPKESVITGHFE 686  
QY 145 GGTFRNDGAVALARSGGYTTDDGYTFNASDI-IEDTGDAIVPHGDHYHIIPKNELAS 203  
DB 687 KEKPKK-----IFDSEPLCLSDN-----LNMQDNF-----DPLWQ 717  
QY 204 ELAAAEAFVLSGRGLNSRTYRONSNDTSTRNWPVSYNFGTT-----NTNTSNNTNS 259  
DB 718 ELSEMFLEQKLNKLSKSEKHIINDLOT-----ELKMGFTFAMLETSCRNSLDL 771  
QY 260 QASQNDIDSLKQLYKPLSQRYVESDG-----LVFDPAQITSRRARGAVAPHGDHYH 313  
DB 772 QFAEKPGSLSL-----QENVSTKGDYDVTGDTLSTLSSQSPVQV----- 816  
QY 314 FIPYSQMSLEFRIARITPLRYRSNHWPDRPSPQPT-----PEP-----SPGP 361  
DB 817 ---PFS--PETETPRRV-----PPDSLPTQGETPCLDVIYVEDCLHODISPA 862  
QY 362 QPAPNLKIDSNSLVSQLVRKVGEGYVEEKGISRYFAKDLPSFTYKNLSKLSKQESV 421  
DB 863 VTPV-VEILSTDARTSHLDR-----SQSPGSEPTL--RLTESDSV 902  
QY 422 -SHLLAKKENAVR---DOEYDKATNLLTEAHKALFYXKGRNSDQALDK----- 469  
DB 903 LADDILASRVSVGSSLPELGGELHNKRPSEDSHSHRLEKMLEAVETLQNSKDAKEA 962  
QY 470 -LBERLNDSEINKKLYVDLFLAFLPIHPRGLGKPNQIETDEDEVRIAQLADKYTTSD 528  
DB 963 GLVGLASDSSTQSLSLEDSLSAPFAPSPS-LETPDS-LETSV----- 1003  
QY 529 GYIFDEHDIISDEGDVAYTPHMGSHWIGKDSLDEKVAQAAYTK-----KGILPSP 583  
DB 1004 ---DVHEALLDSLSGH-TP-----OKLVPRPKPADSGYETVLESPETLHPAP 1048  
QY 584 DADYKANP--TGDSAAATYNNKVEKERTPLVRLPYMWHTYEVKNGNLIIPKHQHYNIK 641  
DB 1049 EGTADSEPATGGDGH-----GLPNPVI-----VISDAGDGHGTEVT----- 1089  
QY 642 FAWEDDHTYKAPNGVYTTLEDLFAITIKYVEHDERPHSNDGMSAENHGLKKHSDENK 701  
DB 1090 -----EFFTAGSGGYSYRD-----SAFSDNDSEPKR-----SEEPG-----TSFSA 1127  
QY 702 NFKADEBEVEETPAPEVYPOVETEKVEAQLKEAF--VLLAKVTSSLSKANPT 751  
DB 1128 LVIVQEOPLRPVLPBQSPAQDSCLFARKSQPDSCSLALHNSSDLELRAT 1179

RESULT 15  
US-09-513-783A-22

Sequence 22, Application US/09513783A  
Patent No. 6416959  
GENERAL INFORMATION:  
APPLICANT: Giuliano, Kenneth A.  
TITLE OF INVENTION: A System for Cell Based Screening  
FILE REFERENCE: 97-022-L1  
CURRENT APPLICATION NUMBER: US/09/513,783A  
CURRENT FILING DATE: 2000-02-25  
NUMBER OF SEQ ID NOS: 180  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 22  
LENGTH: 1610  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-513-783A-22

Query Match 3.3%; Score 137.5; DB 4; Length 1610;  
Best Local Similarity 19.8%; Pred. No. 0.11; Mismatches 337; Indels 265; Gaps 41;  
Matches 174; Conservative 103;

QY 53 ITDQY-----VTSQDHYHYNGKVPYDAIISELLMKDP-NYKLKDE-----DIV 98  
DB 62 VTTGCGYGLQCFARPRDHMKQIDFFKSMPEGEYQDERTFFQDDGNYKTRAEYKREGDTLY 121  
QY 99 NEVK-GGYVIRVDG-----KYYVYLKDAHADNVKTEINROKOEHS---Q 141  
DB 122 NRIELKIDFKEGDNILGHKLEUYNNSHNYIMADKQKNGIKVNEKIRHNIJEDSGVQLAD 181  
QY 142 HREGTPNDGAV-----ALARSQ-----RYTTDDG-----IFNAS 174  
DB 182 HYQATPTIGDGPVLLPDNHYLSYQSALSCKDNEKRDHWLLEFYAAGITLGMDELTKG 241  
QY 175 DIIEDTGDAIVPHGDHYHIP--KNELASASELAAEA-----PLS 213  
DB 242 DEVGMADLSLV---DALTEPPEIEGEIKRDPMALEAFYDDIVGCTVEKTERIFPLD 298  
QY 214 GRGLNSRTYRONSNDTSTRNWPVSYNFGTTNTSNNSNTNSQASQSDIDSLKQ 273  
DB 299 GDEKGTNSSESKKKCLD-TSQVEGIPS--SKP-TLLANDHGMENNTAGSPFD--FLEE 352  
QY 274 LYKLP---LSQRHVESDGLVDPDAQITSRARGAVAPRGDHYHIFIPYSQMSLEFRIARI 330  
DB 353 RVDYPRDIQSSQNMPEBDASFCFQPOQVLDTDQ---AEPNEH-----RDDGLADL 398  
QY 331 IPLRYRSNHWPDRSPQPTPEPSPQPPAPNLKIDSNSLVSQLVRKVGEGYVE 390  
DB 399 LFVSSGPTNASAFTERDNPS-----EDSYGMLPCDSF---ASTAVVSGEMSVGAPNSPCS 450  
QY 391 EKGISRYFAKDLSEYVKNLE---SKLSKQESVHTTLAKKENAVAPROEYDKAYN 445  
DB 451 ESCVS-----PEVETLEQAPATELSKAAEVESVKQOLPAK----- 485  
QY 446 LITEAHKALFYXKGRNSDQALDKLERLND-----ESNKKKIVDDLFLAFLPIHPRRL 501  
DB 486 -----ALETMADQTTDVVHSPSDTTPRGPTAEALAKLE----- 520  
QY 502 GKPNQIETDEDEVRIAQLADKYTTSDGYIFDEHDIISDEGDV----- 545  
DB 521 -----EITRPVILANVTPSTESDQFLAQMDELLGTAEAHANNIILPTPEDESSTK 573  
QY 546 -VTPHMGSHWIGKDSLDEKVA-----QATYKEGILPSPDADYKANPT 592  
DB 574 DVAPRMEEEIIPGNDTSPKTEETTLPIKMDLAPPEVILLKTELEAP---AKGWSLS 629  
QY 593 GDSAAATYNNKVEKERTPLVRLPYMWHTYEVKNGNLIIPKHQHYNIKFAFMDHDTYKA 652  
DB 630 EIEELAKNDYKSAE-IPVAQ-ETVVSSETVYVLADEVLP----- 667  
QY 653 PNGYTLDELFAITIKYVEHP--DERPHSNDGMGN-ASEHVLGKKD-HSEDPNKNFKADEE 708

```
Db 668 -----SDPTTFLTKDVTLPLEAERPLVDMPPSLETEMILGKETAPPTETNIGMAKDMS 721
QY 709 PVEETPAE-----PEVPQVETEKVEAQLEAEVYLLAKVTDDSSIKANATEETLAGLRNN 760
Db 722 FLPESEVITIGKDVVILPETKVAEFNNV--TPLSEEV--TSVKDMSPSA--ETEAFLAKN 775
QY 761 LTLQ-----IMDNNSIMAEAEKLLALLKGSNPSVSKEK 794
Db 776 ADLHSGTELIVDNS--MAPASDLALPLETKVATVPIDK 812
```

Search completed: November 13, 2002, 04:13:25  
Job time : 43 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 15:30:59 ; Search time 47 Seconds  
(without alignments)  
255.074 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKNRVS...KILALKGNSPSYSKEKIN 796

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCRUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	US-09-765-272-56	Sequence 56, App1
2	2649.5	63.6	763	US-09-765-272-66	Sequence 66, App1
3	1187.5	28.5	447	US-09-765-272-182	Sequence 182, App
4	170	4.1	2478	US-09-815-242-5816	Sequence 5816, Ap
5	170	4.1	2478	US-09-815-242-12967	Sequence 12967, A
6	168.5	4.0	1043	US-09-946-805-4	Sequence 4, App11
7	159	3.8	870	US-09-815-242-5493	Sequence 5493, Ap
8	159	3.8	870	US-09-815-242-12637	Sequence 12637, A
9	151.5	3.6	2086	US-09-815-242-5639	Sequence 5639, Ap
10	151.5	3.6	5795	US-09-815-242-12610	Sequence 12610, A
11	145.5	3.5	2368	US-09-815-242-5635	Sequence 5635, Ap
12	145.5	3.5	2368	US-09-815-242-12369	Sequence 12369, A
13	144	3.5	1040	US-09-765-272-118	Sequence 118, App
14	143	3.4	2025	US-09-815-242-5703	Sequence 5703, App
15	143	3.4	3158	US-09-815-242-12611	Sequence 12611, A
16	142.5	3.4	6281	US-09-815-242-12996	Sequence 12996, A
17	141.5	3.4	2437	US-09-815-242-5834	Sequence 5834, Ap
18	141	3.4	2431	US-09-901-106-2	Sequence 2, App11
19	140	3.4	1111	US-09-815-242-12955	Sequence 12955, A

20	139.5	3.3	1609	US-09-938-275-11	Sequence 11, App1
21	138.5	3.3	1125	US-09-974-298-114	Sequence 114, App
22	138.5	3.3	1349	US-09-815-242-5898	Sequence 5898, Ap
23	138.5	3.3	1349	US-09-815-242-13137	Sequence 13137, A
24	138.5	3.3	1703	US-09-801-368-340	Sequence 340, App
25	138	3.3	1029	US-09-815-242-5885	Sequence 5885, Ap
26	138	3.3	1048	US-09-815-242-13083	Sequence 13083, A
27	137	3.3	1016	US-09-815-242-5845	Sequence 5845, Ap
28	135.5	3.3	2076	US-09-815-242-5815	Sequence 5815, Ap
29	135.5	3.3	2165	US-09-815-242-12913	Sequence 12913, A
30	135	3.2	2665	US-09-864-761-34248	Sequence 34248, A
31	134	3.2	1163	US-09-893-348-18	Sequence 18, App1
32	134	3.2	2308	US-10-000-954-2	Sequence 2, App11
33	133.5	3.2	1607	US-09-938-275-10	Sequence 10, App1
34	132	3.2	1164	US-09-950-0464-2	Sequence 2, App11
35	131	3.1	915	US-09-817-514A-6	Sequence 6, App11
36	130.5	3.1	704	US-09-801-368-218	Sequence 218, App
37	130.5	3.1	2434	US-09-815-242-5835	Sequence 5835, Ap
38	128	3.1	1230	US-09-727-384-8	Sequence 8, App11
39	127	3.0	1018	US-09-815-242-5797	Sequence 5797, Ap
40	127	3.0	1018	US-09-815-242-12838	Sequence 12838, A
41	126	3.0	1202	US-09-864-761-43061	Sequence 43061, A
42	124.5	3.0	789	US-09-995-587A-1	Sequence 1, App11
43	124	3.0	1501	US-09-924-154-17	Sequence 17, App1
44	123.5	3.0	1601	US-09-862-027-40	Sequence 40, App1
45	123.5	3.0	1781	US-09-738-877-3	Sequence 3, App11

ALIGNMENTS

RESULT 1  
US-09-765-272-56  
Sequence 56, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 796 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 56:

US-09-765-272-56

Query Match	100.0%;	Score 4163;	DB 10;	Length 796;
Best Local Similarity	100.0%;	Pred. No. 1.5e-265;		
Matches 796; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy	1	SYELGLYOARFKENNRASYIDGKOATOKTEMLTDEVEKREGINAEOQVITITPOGYAT	60
Dp	1	SYELGLYOARFKENNRASYIDGKOATOKTEMLTDEVEKREGINAEOQVITITPOGYAT	60
Qy	61	SHGDHYHYNGKVPYDAIISELLKKDPNYKLAKDEDIVNEVGEVGYIKYDGKYYVYLKKA	120
Dp	61	SHGDHYHYNGKVPYDAIISELLKKDPNYKLAKDEDIVNEVGEVGYIKYDGKYYVYLKKA	120
Qy	121	AHADVNRKEELNROKQEHSHQREBGTPRNDCAVALANSQGYTTDDGYIFNASDIIEBT	180
Dp	121	AHADVNRKEELNROKQEHSHQREBGTPRNDCAVALANSQGYTTDDGYIFNASDIIEBT	180
Qy	181	GDAYIVPHGDHNYIPKNLSASELAAAEAFISGNGNLSNSTYRPRONSDNTRTMWVS	240
Dp	181	GDAYIVPHGDHNYIPKNLSASELAAAEAFISGNGNLSNSTYRPRONSDNTRTMWVS	240
Qy	241	VSNPCTTNTNNTSNNTNSQASQSDNDISLQOLKPLPSQHNVESDGLVFPQAOTTSRT	300
Dp	241	VSNPCTTNTNNTSNNTNSQASQSDNDISLQOLKPLPSQHNVESDGLVFPQAOTTSRT	300
Qy	301	ARGVAVPHGDHYHETPYQSOMSELEERLARIPLRYRSNHWVPDSRPEOPSPOTPEPSFG	360
Dp	301	ARGVAVPHGDHYHETPYQSOMSELEERLARIPLRYRSNHWVPDSRPEOPSPOTPEPSFG	360
Qy	361	POPANLNIDSNSLSVSOLVKRVGEGYVPEEKGISRYVPAKDLPEETVYNLESKTSKOPS	420
Dp	361	POPANLNIDSNSLSVSOLVKRVGEGYVPEEKGISRYVPAKDLPEETVYNLESKTSKOPS	420
Qy	421	VSHTLTAKKENYAPRDOEFTYDKAYNMLTEAHKALPEXNGRNSDFOALDKILERLNDESTN	480
Dp	421	VSHTLTAKKENYAPRDOEFTYDKAYNMLTEAHKALPEXNGRNSDFOALDKILERLNDESTN	480
Qy	481	KEKLVDDLLAPLAPITTHEBRIGKPNISOIEFTYDEVRIAQLADKYTTSDGYITDEHDIISD	540
Dp	481	KEKLVDDLLAPLAPITTHEBRIGKPNISOIEFTYDEVRIAQLADKYTTSDGYITDEHDIISD	540
Qy	541	EGDAVTPHMGSHWIGKDSLSDEKRYVAAQATYKKGILPPSPDADVYKANPTGDSAAATY	600
Dp	541	EGDAVTPHMGSHWIGKDSLSDEKRYVAAQATYKKGILPPSPDADVYKANPTGDSAAATY	600
Qy	601	NRVKGEKRIPLVRLPYMEHTVEYKNGNLIIPKHDIYHNIKFAMFEDDHTYKAPNGYTTLED	660
Dp	601	NRVKGEKRIPLVRLPYMEHTVEYKNGNLIIPKHDIYHNIKFAMFEDDHTYKAPNGYTTLED	660
Qy	661	LEATIKYVVEHDEPRHSDNGMNASERYVLGKKHSDSDPNKFKADEVEETPAPEVP	720
Dp	661	LEATIKYVVEHDEPRHSDNGMNASERYVLGKKHSDSDPNKFKADEVEETPAPEVP	720
Qy	721	QVETEKRYVAQLKEAVLLAKYVDSLSKANATETLAGLRNNLTLOIMDNNSIIMAEAEKLLA	780
Dp	721	QVETEKRYVAQLKEAVLLAKYVDSLSKANATETLAGLRNNLTLOIMDNNSIIMAEAEKLLA	780
Qy	781	LIKGSNPSSVSKKKTIN 796	
Dp	781	LIKGSNPSSVSKKKTIN 796	

RESULT 2  
 US-09-765-272-66  
 ; Sequence 66, Application US/09765272  
 ; Patent No. US20020061545A1  
 GENERAL INFORMATION:  
 ;  
 ; APPLICANT: Choi et. al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 ; NUMBER OF SEQUENCES: 452  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

CITY: ROCKVILLE

STATE: Maryland  
COUNTRY: USA

COUNTRY: USA  
STD: 30950

ZIP: 20830  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MS

```
OPERATING SYSTEM: MS
SOFTWARE: ASCII Text
```

```

; SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
;

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: IIS/00

```

APPLICATION NUMBER: US/0  
FILING DATE: 22-10-2001

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

APPLICATION NUMBER: 06/301,063  
FILING DATE: <unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Brooks A Anders

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 66:

; INFORMATION FOR SEQ ID NO: 66:

; SEQUENCE CHARACTERISTICS:

LENGTH: 763 amino acids

TYPE: amino acid

STRANDEDNESS: single

```

;      TOPOLOGY: linear
;
;      MOLECULE TYPE: protein
;
;      SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-372--66

```

```

Query Match          63.6%; Score 2649.5; DB 10; Length 763;
Best Local Similarity 66.8%; Pred. No. 2,96-166;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;

QY      1 SYELGLVOARLV-KENNRYSTIDGKQATQKTENITPPEVSKREGINAEQVIYIKTTDGYV 59
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2 SYELGRHQAGGVKKESNSVSYIDGDAQQKEMNITPPEVSKREGINAEQVIYIKTTDGYV 61
QY      60 TSHGDHYHYNGKPYDAIIEELLKMPNKLDEDEIVNFKVGAYKVKVGKYYVYLKD 119
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      62 TSHGDHYHYNGKPYDAIIEELLKMPNQLDSDIVNETKGGIYKVGKYYVYLKD 121
QY      120 AAHADNVATKEEINRQKQESHQREHGGTPRNDGAVAIARSGRYTTDDGYIFNMSDIED 179
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      122 AAHADNITKEETKEIKRQKQESHNNHNS--RADNAVAAARAAGRYTTDDGYIFNMSDIED 178
QY      180 TGDYIVPHGHYHYIIPKNELASASELAAAEVFLSGRGLNSRTRYRONSNTSRTNMPV 239
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      179 TGDYIVPHGHYHYIIPKNELASASELAAAEVYMG-----KQGSRPSSSSSYNA 227
QY      240 SVSNPGTNTNTSNNSNTNSQASQNDISLKLQYLPKLSORHVESGLVFPDAQITSR 299
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      228 NPAPRPRISENNILVTPIYH--NOGENTSSILRELIAKPLSERVYESDGLTFDPAQITSR 286
QY      300 TARGVAAPHGDHYHPIFYSONSELEERTARIIPLRYSNMHWPDSPREPQSPQPTPEPSP 359
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      287 TARGVAAPHGHNHYHPIFYEQNSELEKRIARIIPLRYSNMHWPDSPREPQSPQPTPEPSP 346
QY      360 GPQAPNLIK-IDSN---SLYSQLVKRYVBEGYVEEKGISRYVYAKDLPSSTVYNLESKL 415
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      347 SPQAPNPQAPNSPIDEKLVKEAVRKYVGDDGYVEEKGVSRYIPAKDLASATAAGIDSKL 406
QY      416 SKQSVSTHITLAKKENVAPRDOEYEDKAYNLITPAHKALFNKGRNSDFQALDLKLERLN 475
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      407 AKQGSLSHIKLGAKKTDLPSSDREFFYNKAYDILAIIHQDLDNKRQYDFEALDMLLERLK 466
QY      476 DESTNKEKLVDDLLAFLAPITHPERLKGPNQSOLEYTEDEVARIQALADKYTTSDGYIFDEH 535
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      467 DVXSDKVKLVDDILAIAPLAPITHPERLKGPNQSOLEYTEDEVARIQALADKYTTSDGYIFDPR 526

```



```

Oy 536 DIISGGAAYTPHNGSHWIKGDSLSPKKEYAAQATYRKGLTPSPADVAKNAPTGDG 595
Db 527 DITSDEGAYTPPHTHSHWIKDSLSEKERAQAATYAKKGLTPSTDHDSGNTKAGK 586
Oy 596 AAAIYNRYKGEKRJPLVLYPWVEHTVEKKNLIIIPKDHYNHNIKFAFDHTYKAPNG 655
Db 587 AEAATNRYKAAKKVPLDMRPYLYQTVKKNGLIIIPHDHYNHNIKFEWFDGVLKAPNG 646
Oy 656 YTLDELATITTYVYHPDEPRHSDNGKGNASHVYCKKDHSDPKNKNKADEE-----P 709
Db 647 YTLDELATITTYVYHPDEPRHSDNGFNASHADYVORNNKGADTQOTEPKSEKQOTEPK 706
Oy 710 VEET-----IAEPPRPOVETEKYEAOLKEAEVLLAKYT 743
Db 707 EEPTREREKPKSEKPKPTPEEPSESESEEPVEYTEKVEKLREADDLGKTID 763

RESULT 3
US-09-765-272-182
: Sequence 182, Application US/09765272
: Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-765-272-182

Query Match 28.5%, Score 1187.5; DB 10; Length 447;
Best Local Similarity 46.3%, Pred.No. 1,1e-70;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10.

Oy 4 LGLIQAIFVYKNNRYSYIDGKQATOKTENLAPDEVSKREGIAEDIVIKITDGYVTSHG 63
Db 1 LNHRSQENKNNRYSYIDGSSQSKSENLPDQYQSKREGIQAEIVIKITDGYVTSHG 60
Oy 64 DHHYVYNNKVYDITISPELLMKDNYTKEDYIENEYKGVYIKVDGKYVYLLDAHA 123
Db 61 DHHYVYNNKVYDITISPELLMKDNYTKEDYIENEYKGVYIKVDGKYVYLLDAHA 120

```

```

OY      124  DNVRKKEINRKKOEHSGRGGPRNDGAVALARSGRTTDDGYIFENASDIIEEDGDA 183
Db      121  DNVRKKEIDNNKKOEHKVDN-----KNSNVAARSGCRTTNGCYFNPADIIEDGNA 176
OY      184  YIVPGDHYHYHPIPNKELASAEELAAEAFISGRGNLSNSRTYRRONSDNTSRTVMVPSVN 243
Db      177  YIVPGDHYHYHPIPSDLSASLAAKAKHLACKNNQPSQLSYSTASDN----- 224
OY      244  PGTNTNTSNNSTNSQASQSGNDIDSLIKOLYKPLSQRHVESDGLVDPQAQITSRTARG 303
Db      225  ---NTQSVAGSGTSKPKNSKSENQSLIKELYPDSPAQRYSSESGLVFDPAKIISRTPNG 280
OY      304  VAVPBGDHYHYHPIPTSONSELEERLARIIPLYKRSNMHVPPSPRPQSPPTPEPSPQPP 363
Db      281  VAVPBGDHYHYHPIPTSKLSALEEKTARMP----- 309
OY      364  APNLKIDSNSSLVOLARKVGEQGVFFPEKIGISRYVFAKDLPSETVKMLSEKLSKQESVSH 423
Db      310  -----ISGTGVSTYSNNK-----PNEVSSLSGLSSMPS--- 339
OY      424  TLTKAKENAVARDFEFDKAVNLLTEAHKALFXMKGRNSDFQALDKLELNDESTYKEK 483
Db      340  -LTTSKELSSASDGYIFNPK-DIYEETATAYVRHG---DHFHVYIPK-----SNQIGQPT 389
OY      484  LVDDLAFIAPLITPERIGKRNQSIETEDEVRIAQLADKYTTSQGYIFPEHDIISDEG 543
Db      390  LPNNSLAPSP-SLPINPGTSHENH-----EDGYGFDPANRIIAEDS 431
OY      544  AYTVPHPHGHSHMIGK 558
Db      432  GFVMSHGDNHNYFPK 446

RESULT 4
US-09-815-242-5816
: Sequence 5816, Application US/09815242
: Patent NO. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsten, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5816
: LENGTH: 2478
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-5816

```

Query Match 4.1%; Score 170; DB 10; Length 2478;  
Best Local Similarity 19.4%; Pred. No. 0.0072;  
Matches 169; Conservative 141; Mismatches 354; Indels 206; Gaps 38;

```
QY 8 QARTVKNRVSIIYDQKQTKT-----ENLTPEVSKREGINAEQIYIKITDGYVTS 61
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 986 QATTTKSKAKKEIAQKASERKTAIEAMNDSTTEEQAAKDKV-----DQAVVTA 1035

QY 62 HGD-----HYHYNGKVPYDAIISELLMKDPNKKLDEDIYNEVKGGVYIKVDGKYV 115
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1036 NMDIDNAANNVDNAKKTNEATIA--ATTPDANKVPAKQAIADKVQAOETAIDG---- 1089

QY 116 YLKDAHADNVPTKEEINRQKOEHSOHRREGTFRNDGA-----VALAR---SOG 161
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1090 -----NNGSTTEEKAAKQOVOTEKTTADAIDAHTNAEVEAKKAIAIKIEAIOP 1141

QY 162 RYTTDDGYIFNASDIIEDTGDAIYVPHGDHYHIFPKNELSASELAAEAFLSGRGMLSNS 221
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1142 ATTTKD-----MKEAIAATKANEKTA-----IAQTODITAEIIAANA-----DVDNA 1185

QY 222 RYTRRONSNTSRNTNVPVSNGTNTNTSN--NSNTNSQASQSDIDSLIKQLYKPL 279
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1186 VT---QANSNIEANSQNDVDQAKTGENSIDQVTPTVKKATARNETIALLNN--KLQE 1240

QY 280 SQRHVESDGLVDPDPAQITSRARGVAVPHGDHYHIFPYSQMSLEERIRARIITPLRYSNH 339
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1241 IQATPRAIDEEKQAADAENKENGKA-----NQAIISAATTNNOVDEAKANAEA 1288

QY 340 WVPDSRPEQSPQPTPEPSPGQPA-----PNLKIDSNSSLVSQLYRKVGEG----- 386
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1289 AINAVTPKVVKKQAQADEIDQATQNTVNIINDQNAITEEKEAIIQOLATATVADAKNNIT 1348

QY 387 YVEEKGISRYVFAKDLPESEVKNLES-----KLSKQESVSHLTAKKENVAPRDGEFYDK 442
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1349 AATDDNGVQO---AKDAGKNSIOSTOPATAVKSNAKNVDQAVTTONQAI-----DN 1397

QY 443 AYNLLTEAHKALFXNKRNSDFOALDKLLER-LNDESTN-----KEKLYDDLAFIAPIT 496
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1398 TTGATTEE-----KNAKADLVIAKAKEKAYODILMAQTNDVYQIKQAVADVIOGITADTT 1452

QY 497 HEBRLCKPMSQIEYTEDEVRIAQLADKITYTSQGYIFDEHDIITSDCEDAVYTPHMGSHMT 556
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1453 IKD-VAKDELATRKANQOKALIQATADATT-----EKEEONNOQVADQILT--QGNQNTIE 1502

QY 557 GKDSLDFKFAVAQAATYKKGILPPSPDADVKANPTGDSAAAIYVNRKGEKRIPIVRLRY 616
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1503 NMSIDVDVTAQDNAL---QALDPIDASTDVKTNNARAELLTEKQNTI-----TEIIN 1551

QY 617 MVEHTVEVKNGLIIP---HKDHYHNIFAFW--DDHTYKAPNGYTLDELPAITIKYVE 670
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1552 NNHTTEER-KNDIGFVRAVEGLNINNAATTTGDTVTAADPAVQKVOQLHA----- 1603

QY 671 HEDERHSDNGMGNASEHVLGKKHSEDPKNKFADEE-PVEETP--APEEVPQVETEYK 727
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1604 NPYKKP-----AGKKEILDQ---AAADKKQIIOGTNPASQOEIINDAKOE-V 1644

QY 728 EQLQLEA-----EVLAKVYDSSLKANATETLAGIRNNLTLOIWD----- 767
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1645 DIELNDAKTNVDOSSITNEYDNAVKEGKAKINVKITFSEYKDALAKIEDAIYAKVNEAD 1704

QY 768 -----NNSIMAEKILLALLKGSNPSVSK 792
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1705 NSMASTSSIEAEKKQALAEIKOTADQNVNQ 1734
```

RESULT 5  
US-09-815-242-12967  
; Sequence 12967, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.

```
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ElitRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 12967  
; LENGTH: 2478  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12967
```

Query Match 4.1%; Score 170; DB 10; Length 2478;  
Best Local Similarity 19.4%; Pred. No. 0.0072;  
Matches 169; Conservative 141; Mismatches 354; Indels 206; Gaps 38;

```
QY 8 QARTVKNRVSIIYDQKQTKT-----ENLTPEVSKREGINAEQIYIKITDGYVTS 61
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 986 QATTTKSKAKKEIAQKASERKTAIEAMNDSTTEEQAAKDKV-----DQAVVTA 1035

QY 62 HGD-----HYHYNGKVPYDAIISELLMKDPNKKLDEDIYNEVKGGVYIKVDGKYV 115
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1036 NMDIDNAANNVDNAKKTNEATIA--ATTPDANKVPAKQAIADKVQAOETAIDG---- 1089

QY 116 YLKDAHADNVPTKEEINRQKOEHSOHRREGTFRNDGA-----VALAR---SOG 161
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1090 -----NNGSTTEEKAAKQOVOTEKTTADAIDAHTNAEVEAKKAIAIKIEAIOP 1141

QY 162 RYTTDDGYIFNASDIIEDTGDAIYVPHGDHYHIFPKNELSASELAAEAFLSGRGMLSNS 221
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1142 ATTTKD-----MKEAIAATKANEKTA-----IAQTODITAEIIAANA-----DVDNA 1185

QY 222 RYTRRONSNTSRNTNVPVSNGTNTNTSN--NSNTNSQASQSDIDSLIKQLYKPL 279
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1186 VT---QANSNIEANSQNDVDQAKTGENSIDQVTPTVKKATARNETIALLNN--KLQE 1240

QY 280 SQRHVESDGLVDPDPAQITSRARGVAVPHGDHYHIFPYSQMSLEERIRARIITPLRYSNH 339
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1241 IQATPRAIDEEKQAADAENKENGKA-----NQAIISAATTNNOVDEAKANAEA 1288

QY 340 WVPDSRPEQSPQPTPEPSPGQPA-----PNLKIDSNSSLVSQLYRKVGEG----- 386
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1289 AINAVTPKVVKKQAQADEIDQATQNTVNIINDQNAITEEKEAIIQOLATATVADAKNNIT 1348

QY 387 YVEEKGISRYVFAKDLPESEVKNLES-----KLSKQESVSHLTAKKENVAPRDGEFYDK 442
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1349 AATDDNGVQO---AKDAGKNSIOSTOPATAVKSNAKNVDQAVTTONQAI-----DN 1397

QY 443 AYNLLTEAHKALFXNKRNSDFOALDKLLER-LNDESTN-----KEKLYDDLAFIAPIT 496
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1398 TTGATTEE-----KNAKADLVIAKAKEKAYODILMAQTNDVYQIKQAVADVIOGITADTT 1452
```

Qy 497 HPERLGRPNQSIETEDEVRIAOQLADKYYTSDGYTFDEHDIISDEGDAYVTPHMGSHWI 556  
Db 1453 IKD-VAKDELATFKANEQKALLAQATADATT-----EKEQANQGVADQOLT--QGNQNIIE 1502  
Qy 557 GKDISDSEKAAQAAYTEKEGILPPSPDADYKANPTGSAAIYRVKGEKRIPLVRLPY 616  
Db 1503 NAQSIDVTAKDNAI---QAIDPIQASTDVKTNAABELTEMQNKI-----TEILN 1551  
Qy 617 MVEHTVEYKNGNLIIIP---HKDHYHNKIFAMF--DGHITYAPNVTYLEDATIKIYVE 670  
Db 1552 NNETTNEK- GNDIGPVAAVEEGININAAATTGGDVTTAKDTAVQKYQOLHA----- 1603  
Qy 671 HDERPHSNDGNASEHVLRKDHSEDPNKNFKADEE-PVEETP--APEVPQVETEKV 727  
Db 1604 NPVKP-----AGKKELDQ-----AADKKTOIEQFPNNAQOEINDAKQE-V 1644  
Qy 728 EAQLKEA-----EVLAKYVDSLSKANATETTLAQRNLTLOQMD----- 767  
Db 1645 DTELQAATNVDOSTNYVNAVEGRAKINAVKTFSEYKDALAKIEDAVNAKVNEAD 1704  
Qy 768 -----NSTMAEAEKLLALLKGSNPSYSVK 792  
Db 1705 NSMASTSEIAEAKOKLAEKOTADQNVNQ 1734

## RESULT 6

US-09-946-805-4  
; Sequence 4, Application US/09946805  
; Patent No. US20020116734A1  
; GENERAL INFORMATION:  
; APPLICANT: Dickman, Martin B.  
; TITLE OF INVENTION: PLANT DERIVED BAG HOMOLOGUES  
; FILE REFERENCE: 480140.469  
; CURRENT APPLICATION NUMBER: US/09/946, 805  
; CURRENT FILING DATE: 2001-09-04  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-946-805-4

Query Match 4.0%; Score 168.5; DB 10; Length 1043;  
Best Local Similarity 18.5%; Pred. No. 0.0026;  
Matches 169; Conservative 108; Mismatches 327; Indels 311; Gaps 36;

Qy 40 KREGINAEQIVIKITDQGYVTHSGDHYHYNGKV--PYDALISELLMKDPNRYLKDED 96  
Db 271 KRQVASE-----SSNEDRKKMQNGKVEYEPFDSIMKSL----- 306  
Qy 97 IVNEKGGVYIKVDGKYVYLKDAHADNVKKEI-----NRQKQESHQ 142  
Db 307 ---IQG-----QDVKEAQNOKNKEPGOVPIPIFWIPSYGKRDVEASES 348  
Qy 143 REGGTP-----RNDGAVVALARSQGRYTTDDGYIFNASDILIBDTGDAYIVPGSD 190  
Db 349 KESSEGNELSCPSDLHRNEQILIQAKGK-----EGNFEQVLSDAEEKSVINIPVAN 403  
Qy 191 HY-----HYIPKNELISASELAAEAFISGR----- 215  
Db 404 HLOEPRIIPVLTSEHNLKRPTEPTRIAKNEPVSTKKEQSSSSSEASKLPPVCLRVPL 463  
Qy 216 -----GLNSKRYTRONSQNT-----SRTNWVPVSNGPTTNTNTS--NN 254  
Db 464 PKERNGKSVSHPRKMSKETKIAPLSSKKAESRT--VPEACNVCEADANEMKMAE 521  
Qy 255 SNTNSQASQNDIDSLTKQLKPLSQRHVESDGLVFPDPAQTTSTARGAVAVPHGDHYHF 314  
Db 522 GSLNALRTEKGSVES-----NSNIQESNGRITIPCKEKKERQD----- 561  
Qy 315 IPIYSOMSELERIARIIPLRYS--NHMPVDSRPEQSPQPTPEPSPG----- 360

Db 562 ---AKKSFTEEEAARIIOSMYRGYDVRWREPIKLLKEIA---TVREQMGDVAKRIEALEA 615  
Qy 361 --PQAPNLIKIDNSNSLSVQYVRKGEVPEEKISIRYVFAKDLPSERYKULESKLS-- 416  
Db 616 STDQHIIEKEIYVNGELVNNLLKLD---AVGCLPSTREFKALATE-LSSIQKRLSL 671  
Qy 417 KOESVHTLTAKKENAVARDQEFYDRAVNL---LTEAHKALFXKKGRNSDQALDKLE 472  
Db 672 KNSCSAKKEAVKEQVELKSP--SDSPVNLHSQLTEENKMW----- 712  
Qy 473 RLNDESTKKEIYDLDLAPLAPITPERLGRKPNISOIETTEDEVRIAOQLADKYYTSDGYIF 532  
Db 713 ---SDTLNLEKYL---LSPEEH-----PMSVYLNFTDEK---QAESAAETEgy-- 751  
Qy 533 DEHDIISDEGDAYVTPHMGSHWICKDSIDSEKVAQAAYT-----KEGILPPSPDA 585  
Db 752 -----GLEFTLATDSKQATENAAASSTIPIPKIGEVETVGNP-- 791  
Qy 586 DYKANPTGDSAAAIYRVKGEKRI-----PLVRLPYMVEHT-----VEV--- 624  
Db 792 -----PSADGNGMTVTYVNEKNAMVYESLEPIINLPPQVETETINSTIDPENASEVSA 846  
Qy 625 -----KGNLIIPKHD---HYHNKIFAMFDHT---YKAPNGYTLIEDLFA--- 663  
Db 847 ETNSSENENRKGEDDIVLHSEKNVELSELPGVIDEETQPLSDQSSSYTRGNMTAMD 906  
Qy 664 -TIKYVHPDERPHSNDGNASEHVLRKDHSEDPNKNFKADEPE-----ETPAE 716  
Db 907 KTAQSETEVHSPNNSGIGQOTSEPODEKQOS--PETEVYVKQPLETEVYILNQAE 964  
Qy 717 PEVPQVETEKYEAQKEAEVLLAKYVDSLSKANATETTLAQRNLTLOIMDNSIMAEE 776  
Db 965 PEITPBGISKETKRIKMEENQRFETIMELVYKAGRDQ--LEVYSKITSYKSLKELKSHK 1022  
Qy 777 KLALLKGSNDSVS 791  
Db 1023 KTOIRRRASRKMVS 1037

## RESULT 7

US-09-815-242-5493  
; Sequence 5493, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlssen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815, 242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 5493
;
; LENGTH: 870
;
; TYPE: prt
;
; ORGANISM: Staphylococcus aureus
;
US-09-815-242-5493

```

Query Match	3.8%	Score 159;	DB 10;	Length 870;
Best Local Similarity	19.3%;	Pred. No. 0.0086;		
Matches 164;	Conservative 120;	Mismatches 273;	Indels 294;	Gaps 45

OY	14	ENNRVSYPIDGQAOIQKLENL-----PDEVSKREGJIN-AEOIV	50
Dd	54	ONNNAFTTTPPKPDINOOTOPATOPANTAKNPRADESCKDAIKNPAYNKHEHDIGPREQVN	113
OY	51	IKITDOGVYTSHGDIHYHNYNGKVRYDAIISEBLLMKDPNYKLKDEDIVNVKGGYVKVD	110
Dd	114	FÖLLDKNNET--QYHFHFISIKDPAUY-----YTKKKAEBELDINNASTWK--	157
OY	111	GKIYYUUKDAAHADNVRTKEBINOK-----OEHSHOREGGTRPNDCAVALAR	158
Dd	158	-KFEEY-----ENNOKLPVRVLSYSPVEDHAZYR---PFVSDGTOL-K	197
OY	159	SÖGRTTDDGY-----IENASDIE-DVGDAIYVPHGDHYIIRKENLSEAS	203
Dd	198	IYSSTOIDDGAEITYDYTKLVFAKPITNDPSLVSDTDNA-----VATNDQSSS	246
OY	204	EIAAAEAFSLSGRNILMSRTYRRONSDNTSRKNVP-----SVSNPG-	245
Dd	247	DAS-----NOTMNT-SNOGSTTNANNQOPATIDMSGPAPQKSSANADAQSSQ	255
OY	246	---TTTTNTSNASNTPNSOASOSANDISLKOQLKPLPSORHVESOGVLFD---PAQTIS	298
Dd	296	PAHEFNSGNNTNDTFNSSNSOD-----YNQOYPPADESLQDAIKNPALIDK	342
OY	299	RTARGVAAPHGDHNFPIPYQSMLEERIAPIPLRASHHWVPDSRPQPPEPS	358
Dd	343	EHT-----ADNRRPIDFOKMNKDGER-----OFYHYA-----STEVEA	375
OY	359	---POPAPAPNLK-----IDSNSLYSQLVRKVGGE---GYV-FEEKGISRYV	398
Dd	376	TVIPTFKPIIELLKATASTWKKREVEEGOKKLPVELVYSVDSOKDAYIIFFPYSNGREV	435
OY	399	PAKDLPESTYVANKLESKUSKOEVSHTLTAKKENAVPRDQEYPD-KAYN--LITEAHKAL	454
Dd	436	---KIYSSIYGENIHEDYDTLMVFAQPIITNPNDDYVDEETYNILOKLAPYHKA-	487
OY	455	FAXNGRNSDFALDKLERLUNDE-STKKEKYUDLLAFLA----PITHBERLGKPRSQI	508
Dd	488	---FTLERQOYELEKLOEKLRPEKKAERYKKKLDOTRVELADQYKSATFEENVTPITDOL	544
OY	509	EYTEDEVENIAOLAADKYTTS--DGYIPDEHIISD--EGDAUVVT-PHMGSHTWGKDSLSD	563
Dd	545	TDVOGAHHVVESESSENSVMDGV- EHPRYTTLTGOKYVIMKPTGDSYW--KDLIIVE	600
OY	564	KEYVAQAQY-TKEKG---ILPPSDADVKANPTGDSAAILYNRKYGSEKRIPLVRLPYME	619
Dd	601	GKRVTYTSKDKTKNNSRTLIFEYIRD-----KAYYNAI--YKVVANIGIEGQ	645
OY	620	HTEVEKNCNLIIPKHDMHNKKFMPPDHYYKAPNGTYLTLEDLFATIKYYUEHNDERHSN	679
Dd	646	YHVRIJN-----QDIIKO--DDDTSQ-----	664
OY	680	DGMGNASEHYGLKDKHSDBPKNKKADEBEVEELTPAPEVPOVETEKEVAOLKEAEYLLA	739
Dd	665	---NNTSEPRLVVQGOGEDBNVYATGTATANNISLETTPSB-----ATDKVD--LLEPSDMV	713
OY	740	KVYDSSLKANA 750	
Dd	714	KDTDSYVDKDA 724	

Sequence 12637, Application US/09815242	
Patent No. US20020061569A1	
GENERAL INFORMATION:	
APPLICANT: Haselbeck, Robert	
APPLICANT: Ohlsen, Kari L.	
APPLICANT: Zyskind, Judith W.	
APPLICANT: Wall, Daniel	
APPLICANT: Trawick, John D.	
APPLICANT: Carr, Grant J.	
APPLICANT: Yamamoto, Robert T.	
TITLE OF INVENTION: Identification of Essential Genes in	
FILE REFERENCE: ELITRA.011A	
CURRENT APPLICATION NUMBER: US/09/815,242	
CURRENT FILING DATE: 2001-03-21	
PRIOR APPLICATION NUMBER: 60/191,078	
PRIOR FILING DATE: 2000-03-21	
PRIOR APPLICATION NUMBER: 60/206,848	
PRIOR FILING DATE: 2000-05-23	
PRIOR APPLICATION NUMBER: 60/207,727	
PRIOR FILING DATE: 2000-05-26	
PRIOR APPLICATION NUMBER: 60/242,578	
PRIOR FILING DATE: 2000-10-23	
PRIOR APPLICATION NUMBER: 60/253,625	
PRIOR FILING DATE: 2000-11-27	
PRIOR APPLICATION NUMBER: 60/257,931	
PRIOR FILING DATE: 2000-12-22	
PRIOR APPLICATION NUMBER: 60/269,308	
PRIOR FILING DATE: 2001-02-16	
NUMBER OF SEQ ID NOS: 14110	
SOFTWARE: seqid for windows version 4.0	
SEQ ID NO: 12637	
LENGTH: 870	
TYPE: PRF	
ORGANISM: Staphylococcus aureus	
US-09-815-242-12637	
Query Match	3.8%; Score 159; DB 10; Length 870;
Best Local Similarity	19.3%; Pred. No. 0.0086;
Matches 164; Conservative 120; Mismatches 273; Indels 294; Gaps	
QY 14 ENNRYSIDKRAKATOKTEULT-----PDEVSKREGIN-AEQIV 50	
DB 54 QNNNATTTQPPPDYOTQTOPANTAKNYPAPADESLKDAIKNPAVENENHDIGPREQVN 113	
QY 51 IKITDOGYTSHGDHNYHNGKVPYDAIISEELMKDPYKKKDEDIVNEVGQYIKVD 110	
DB 114 FOLLDKNNET---QYHFHSIKDPADV-----YTKKKAVELDINTASTWK-- 157	
QY 111 GRIYYILKDAHADNVKKEETNRQK-----QESHQHRBGCTPRNDGAVALAR 158	
DB 158 -KFEYV-----ENNOKLPVRLVSYPVEDHAYIR--FVPSDQIOEL-K 197	
QY 159 SOGRRTTDDGY-----IFNASDIIE-DTGAYIVPHGDHNYIKPKNELSAS 203	
DB 198 IYSSQIIDDGAETINDYTLVFAKPIYNDPSLVKSDTND-----VATNDQSS 246	
QY 204 ELAAAEAFELSGRGLNSNSTRYRRQNSDSTRTNWP---SVSNPG----- 245	
DB 247 DAS-----NOTNTYT-SNONTSTNNANNOPOATTTDSOPAPQKSSANADOASSQ 295	
QY 246 ---TNTNTNSNSNNSOASQSDNIDSLKOLYKLPLOSRAHESQGLVD---PAQITS 298	
DB 296 PAHETNSNNTDKTNESNOSD-----VNQOYPPADESLDADAIKNPAIDK 342	
QY 299 RTARGVAVPHGDHYFIPIYSQMSLEERLARIILPLKRSNMHWPDSPREOPSPPTPEPS 358	
DB 343 EHT-----ADNMWRPIDQMKNDGER-----QYHHA-----SVPEBA 375	
QY 359 -----PGOPAPNLK-----IDSNSLVSQLYRKKVE---GIV-FEEKGISRYV 398	
DB 376 TIVETKGTGILIGLGTASTWKKEFEYEDCKKLPVELVSYDSDKQYAIRFVPSNGTREV 435	



```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12610
; LENGTH: 5795
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12610

```

```

Query Match          3.6%; Score 151.5; DB 10; Length 5795;
Best Local Similarity 18.5%; Pred. No. 0.39; Indels 253; Gaps 40;
Matches 174; Conservative 133; Mismatches 381;

```

```

QY 1 SYELGLYQARTYKEN-----NRVYIDGKOATQKTENTLPPEVSKREGI-----NA 46
DB 2680 SOATNLAAYESKQSANSLDGMNGNQTALINDKSGTSLASQNLDEQCRNNAYNOAVSNA 2739
QY 47 EGVIVIKITDQGVYTHSGDHYHYNGKVPYDAITISELLMKDPN---YKIKDEDIYNEVGK 103
DB 2740 ETLNKQGT-----PNTAKTAVEQALNNVNSAKHALNGTQNNNAKQ 2781
QY 104 GVIKVDG-----KYYVYLKDDAH-ADNVRTKEELNRKQKSHSOHREGGTPRN---DGA 153
DB 2782 AAITAINGASDLNOKOKALKAQANGAQRVNSAODVQRNATE--LNTAMGTILKHAIDXT 2839
QY 154 VALARS-----QGRYTTDDGYIFNADLIEDTGDAYIVPHGDHYHYIFKNELASAE 204
DB 2840 NTLASKYVNADSTQKNATTK---VTNAEHILSGPTVVTTP-----SE 2881
QY 205 L-AAEAFLISGCGNLSNSETTY--RONSNDNISRTKMWPSVSNPGTTNTSNNSTNSQA 261
DB 2882 VTAAANOVSAAKQOELNGERLERLVAKAQ-----NTAIDATQLTLPQAKALKEQY 2931
QY 266 SOSNDIDSLKQLYKLPISORHVESDGLVDFPAQ---ITSRTARGAAYVPHGDHYHYIFPYQ 319
DB 2932 GQANLLEDT-----QTVOTNGQALNNAMKGLRDSINETTVKASQYNTDASPNN 2980
QY 320 MSELERLARIILPLYRSNHNWVDSRPEQSPQPTPEPSPQAPNLKIDSNS-----373
DB 2981 QSTYSAVSANAGIINTQNTPTMDTSAITQATQYVNNAKNGLNGAENLRNAQNTAKQNLN 3040
QY 374 -----SLVSQLVRRVGEVGYFEENGISRYVFAKDPSE---TYKNLESKLSKQES 420
DB 3041 TLSHLTNMOKSAISSOIDRAGH-----VSEYTAAKNAATELMTQNGNLEQALHDQNT 3092
QY 421 VSHITLAKKENAPRDOEYDKAY-NLITFAIKALFXNKGKRSDDQALDKLLER-----473
DB 3093 VKQGVNFTDADAKRD-----AYTNVARSRAETILNKTOGANTSQODVEAIAIONVTSAKN 3146

```

```

QY 474 -LNDES--TNKRELYDDLAFIAPITTPERLGRPN-----SOLEYTEDEVRIAQ 519
DB 3147 ALNGQONTNNAKNTAKHALNMLTSLINNAQKRDLTTRKIDQATVAGVBAVSNGTQQLNAM 3206
QY 520 -----LADKYTT--SDGIYFDEHDIISDEGDAIYVTPHKGSHWIKGKLSKEKYAAO- 570
DB 3207 ANLQNGINDKANTLASENY-----HDADSDKRTAYTOAVTNAENIINKSGSNLDRKAIVEN 3262
QY 571 ---AATKKEGII-PPS---PDADVKNAPGTDSAAATYNNRKGSKRIPLYLVPVHEVTIVP 624
DB 3263 ALSQVTNAKGAALNGHNLNLEQAKSNANTT-----INGLOHLTTAOKDKLKOQVOQA 3312
QY 625 KNGNLIIPKRDHYHNKIFAMFDDHTYKAPNGYTLLEDFAITYYE-----HPDE 674
DB 3313 QN-----VAGVD--TVKS-SANTLNGAMGTLRNSIQDNTATNNGQNYLDA 3354
QY 675 RPHSDNGMGNASEHYLGRKDHSEDPKKNKA-----DEEPEETPAPPEVQYET 724
DB 3355 TESNKTNNNAVDASANGVYINATSNPMDANAINQIATQVTSKRNALDGTHTNLTOAKQAT 3414
QY 725 EKVEAQLK-----EAEVILAKVTDSILKANATEFLAGLRNLT-----LO--INDNNS--- 770
DB 3415 NALDQATNLNKAKQKALKAQVTSAQRVANVT--SIQQTANELTAMGQLQHGIDDERATQ 3473
QY 771 -----IMAEKELALILKSGNPSVSKER 794
DB 3474 TQKRYRDAQSKKTAYDQAVAAKAILNKQTSNSDKAAVDR 3514

```

```

RESULT 11
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5635

```

```

Query Match          3.5%; Score 145.5; DB 10; Length 2368;
Best Local Similarity 18.2%; Pred. No. 0.27;
Matches 156; Conservative 122; Mismatches 285; Indels 295; Gaps 34;

```



Db 1396 ARDNAL---QALDPQASTVDTKTNARALLTEMQNKI-----TELLNNNETTEER- 1443  
QY 627 GNLIIIP-----HKDHYNIKEAFMDHTYKAPNGYLTLEDLFAITIKYVEHPDERPHSDQM 682  
Db 1444 GNDICGVRAIYEEGLNNT-----NBAITTGVTAKDTAVOKVQOHLANP-- 1488  
QY 683 GNASEHVLCRKHSDPNKFRADDE-PVEETP--AEPEVPOVETEKEVAOLKEAEVLLA 739  
Db 1489 -----VKRPAKGTALDOAAADKKTQIEQPPNASSQOEINDAKOE-VDTELMOAKTIND 1539  
QY 740 K-----VTDSSLKANATETTLAGLRNNLTLOI-----MDN-----NSIMAE 774  
Db 1540 QSTDEYVDNAKKEGKAKINAVKTFSEYKRKDALAKIEAAVNAKVTEADNSNASTSEIAE 1599  
QY 775 AEKILALLKGNPSSVSK 792  
Db 1600 AKOKLAELKOTADQNVNQ 1617  
RESULT 13  
US-09-765-272-118  
Sequence 118, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1040 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 118:  
US-09-765-272-118  
Query Match 3.5%; Score 144; DB 10; Length 1040;  
Best Local Similarity 19.2%; Pred No 0.11;  
Matches 154; Conservative 111; Mismatches 284; Indels 252; Gaps 38;  
QY 107 IKVNGKYV-VLKDAHADNVRTKEINRQKSHQREGGTPRND-----G 152  
Db 76 LKIEGQYIGYIK-TKKQNTELSRVDKYSQNRSGNSTKTSVDYVHASADLEWNGGG 134  
QY 153 AVALARSQGRYTTDDGIYNASDIIEDTGDANYIVPHGDHYHIYIPKNELASASLAAAEFL 212

Db 135 KVSLL---QGEASGDG-----LSEKSSIAADNLSSNDSPA 166  
QY 213 SGRGNLSNRTRYRPNQSDNTSRTNVPVSYPNGPTNTNTSNNSTNSQASQNSDILSLK 272  
Db 167 S-----QVEQNPBHGSESVYRPTYPEQGNFVSATTVQSABEEVLATIND-----R 211  
QY 273 QLYKPLPSGRHVESDG-----LVF-----DPAQITS 298  
Db 212 PEYKPLFETKGTQEPHGEAEAVREDLPVYTKPLETKGTQPGHGEAEAVREEDPAYTEP 271  
QY 299 RTARGVAVPHGHYHFIPIQSOMSELEERLARIPLRYSNMHNVPPSRPQSPQTPRES 358  
Db 272 LATKGTQEPGHE-----GKATVREETLEYTEPVAATKGT-----QEPHEGERVVEEL 319  
QY 359 PGPOPA-----PNL-----KIDNSLSIVOLVR--KVEGYVEE 391  
Db 320 PALEYTRNRKTELQNPYTEETIOPDTLLKNRKRIERQOAGTRITQYEDYIVNGNVET 379  
QY 392 KGISRYVFAKDLPSETVKNLESLSKQESVSHTLAKKENVAPRQEFYDAKYNLL--TE 449  
Db 380 KEYSRTEVA--PVNEVYVGTIVKVPVEITNLTKVEN-----KKSITVSYNLDTTS 431  
QY 450 AHKALEFXNKGSRSDQALDKLERLNDESTNKEKLVDDLALFAP--TH-PELGRPN 505  
Db 432 AYVS-----AKTQVPHG-DKLKVEVDIENPAKEQVYISG-LDYTTYTYVTKTLTYNLGENN 484  
QY 506 SLEIYTEDEVRILAOLADKTYTSDGYTFDEHDILISPGDAYVPHMGHSHWIKSDISKE 565  
Db 485 E--ENTETISTQPOL--EKKIEKIDIDSEVLXGKENDRY-----RXYL--SLSEAP 530  
QY 566 KYAAQAYTKKGIILPPSPDADYKANFTGDSAAIYNRVKGKRIPLVRLPYWEHT--- 621  
Db 531 TDTAKYFVKV-----SDRFKEMYLVPVS-----ITENTDGY 563  
QY 622 -VEVKNGNLIIFHKDHYNIKFAWEDHTY-----KAPNGYT-LEDLFAIK-----Y 667  
Db 564 KYTVAVDQLEVEGTQYK-----DDYTFYVAKSKAEQPGVYTSFKQLVYTMQNSLGV 616  
QY 668 YVEHPDERPHSDGNGNASEHYLAKKQHS-----EDPKNFKADE--EPVEET 713  
Db 617 YTLASDM-----TADVSLGDKQTSYLTGATGSLGSDGKRSATYILKPKPLDPT 667  
QY 714 PAPEVPOVETEKEVAOLKEAEVLLAKYTDSSL-----KANATETLAGLRNNLTLOI 765  
Db 668 LMGATVRRDDITVTSADSKENVAALAKANSANINNVAVEGRISGAKSVAGIASATVTV 727  
QY 766 MDNNSIMAEAEKLLALKGSN 786  
Db 728 IENSSFTG---KLIANHODSN 745  
RESULT 14  
US-09-815-242-5703  
Sequence 5703, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848





Db 1542 YQAAVQ-----HAKDLINQTSNPTL-----DKAQVEQLT 1570  
QY 213 SG-----RGNLSNSRTRYBRQNSDNTSRTNVPSVNGTNTNUS--NNSNTNSQASQ---- 263  
Db 1571 QGVNNAKNDLHGDQKADDDQIAVTDLNLQNSLNNPQRALESQIINNAATREDEVAQKLA 1630  
QY 264 SMDIDSLKQLYKLPISQRHVESDGLVEDPAQITSTARGAVPHGDHY-----HFIPY 317  
Db 1631 AQALDQAMQALRNSIQDQOQTES--GSKFINEDKPOKDAVQAQAAVQAKDLINQTSNPTLDK 1689  
QY 318 SQMSLEERIAIILPIRYSNHWPDSRPEQSPQPTPEPSPGPQAPNKLIDS--NSSL 375  
Db 1690 SQVEQLTQAVTTAKNDLHGDQKADDDQIAVTTVALPMLNHAQOQALTDALINAAPTRE 1749  
QY 376 VQQLVRKVEGVFPEEKGISRYVFAKDLPEFTVKNLESKLS--KQESVHTLTAKENV 432  
Db 1750 VAQHQTATE-----LDHMETLKNKVQOVNTDKAQPNTTEASTDKKEAV 1794  
QY 433 APRDOEYDKAYNLLTEAHKALFXNKGNSDPQALDKLERLND-----ESTNKEKL 484  
Db 1795 ---DQAL--QAESITDPTNG--SNANKDAVEQALTKLOEKYNELNGNERVAEAKAQAKQ 1847  
QY 485 VDDLAFLAPITHPERLGRPNQIETEDVEVRIQDLADKTYTSDGIIFDEHDITISDEGDA 544  
Db 1848 TTDOLAHL---NADQIATRAKONIDQATKLOPIAELVDQATQLONSMDQLQQAVERNANV 1903  
QY 545 YVTPHMGSHWIGKDSLSDEKVEAAQAYTKKEGILPPSPDADVKANPTGDSA--AAIYN-- 601  
Db 1904 EOTVDTYQAD-----SDKQNAVQALAEAEVNL-----KONSNKQOVDQALQNTL 1948  
QY 602 ---RVKGEKRIPLVRLPYWVEHTEVKNGLIIPKDHYNIKPFWEDDHTYKAPNGYT 657  
Db 1949 NAKQALNGDEVALAK-----TNGKHD I--DQNLNLNNAQODGFKGRIDQSHD 1994  
QY 658 LEDLPAITKIYYEHDPDERHSDGNGNSEHYLGKKHSEDPKNKFKADEEPVEETPAEP 717  
Db 1995 LNOIQOIY-----DEAKALNRAMDQLSQETISGNEGRTKGSTNYVNAD-----T 2037  
QY 718 EYPOVETEKEVEAQLKEAEVYLAKVTDSILKA-----NATETLAQLRNN 760  
Db 2038 QVKQYDEAVD---KAKQALDKSTGQNTAEQVYIKLNDVYTAAKKALNGEERLNN-RKS 2092  
QY 761 LTLQIMDNNSIMAEAEKLLALIKSGSNPSSVSK 792  
Db 2093 EALQRLDQTLHLNNAQROLAIQIINNAETLNK 2124

Search completed: November 13, 2002, 04:12:37  
Job time : 63 secs